The Dissertation Committee for Jinseok Shin Certifies that this is the approved version of the following Dissertation:

Essays on Human Capital and Economics of Education

Committee:

____
Leigh Linden, Supervisor

____
Sandra Black Youngblood

____
Daniel Hamermesh

____
Richard Murphy

Essays on Human Capital and Economics of Education

by

Jinseok Shin

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2018

Acknowledgements

I would like to thank Leigh Linden, Sandra Black, Daniel Hamermesh, Richard Murphy, and Sukjin Han for their invaluable guidance and support. I would also like to thank Marika Cabral, Mike Geruso, Stephen Trejo, Haiqing Xu, and the seminar participants at the University of Texas at Austin for their comments and helpful feedback. Special thanks to Vivian Goldman-Leffler and my family.

Abstract

Essays on Human Capital and Economics of Education

Jinseok Shin, Ph.D.

The University of Texas at Austin, 2018

Supervisor: Leigh Linden

Since the publication of the Coleman Report in 1966, economists have contributed significantly to the body of research focused on understanding how school systems work. By approaching the development of students' human capital as a production process in which school resources, teachers, etc., are the inputs, we can utilize the rich empirical and theoretical tool kit developed to understand the production process of firms. However, recent studies show that there are more factors that have not been studied yet. Moreover, not all already known factors have studied thoroughly. My dissertation studies the following questions in an education production context: First, does alleviating adolescents' sleep deprivation improve academic performance? Second, do peer effects affect student performance even across grade? Third, is there a benefit to taking more tests?

First chapter examines how the delayed school start time affects students' sleep schedule, their time use on other activities as well as academic performance. I exploit a natural experiment that delayed school start times in a large province in South Korea using both the Korean Time Use Survey and school administrative data in a difference-in-differences framework. I find that teenagers and middle school students slept an additional 30 minutes per school night by spending less time studying independently. Despite the

reduced study time, I find that delaying the school day in middle school start time by forty minutes increased math test scores by 0.034 standard deviations and reading test scores by 0.022 standard deviations. This confirms that reduced sleep deprivation is the main mechanism of improved academic performance.

The second chapter examines how peer effects work within high school. Most studies on secondary education peer effects have focused on peer effects within a classroom or a grade, but students' peer networks can be much larger, even across grade. Thus, I examine, in work with Jungmin Lee[1], whether students do better when the students in other grades are better. We find no effect of having more high-achieving students in another grade on your own performance, whereas we do find that having more low-achieving students has a negative impact on your performance.

The third chapter examines the impact of eliminating two exams in one year on later test scores by exploiting a phased-in policy in Korea in which students are exempt from second-semester exams in their first year in middle school with a difference-in-differences strategy. I find that taking two tests less during the last school year of middle school decreased math test scores by 0.053 standard deviations, but does not have an impact on reading test scores. This negative impact on math test scores remains relatively constant for at least one academic year. This means more tests can incentivize students to spend additional time studying and in learning more knowledge or test taking skills. However, it is notable that less tests do not have a harmful effect on reading, but on math. This suggests the number of tests has an impact on students' academic performance through learning knowledge than test taking skills since math requires more prior knowledge and is especially learned in class. Moreover, it is less likely that required test

[1] Seoul National University, South Korea, and IZA, Germany

taking skills are different based on subject as most portion of the tests in the data is multiple choice questions.

Table of Contents

List of Tables

List of Figures

Chapter 1: Sleep More, Study Less? The Impact of Delayed School Start Time on Sleep and Academic Performance

1 INTRODUCTION

Adolescents are severely sleep deprived. Pediatricians recommend teenagers get 8-10 hours of sleep per night (Hirshkowitz et al. 2015); however, teenagers in the United States average 7.6 hours (National Sleep Foundation 2006).[2] The main reason is a combination of biological factors faced during adolescence that delay the secretion of melatonin, a hormone that induces sleep, and early school start times (Andrade et al. 1993; Cardinali 2008; Carskadon, Vieira, and Acebo 1993; Carskadon et al. 1997; Crowley, Acebo, and Carskadon 2007; Eaton et al. 2010). Relative to the summer, adolescents lose as much as 2 hours of sleep every night during the school year (Hansen et al. 2005). This may play an important role in student performance. Sleep provides a critical period for the brain to consolidate new information into long-term memories (Fogel and Smith 2011; Gais and Born 2004; Nader and Smith 2003; Stickgold and Walker 2007). Moreover, during the day, sleep deprivation reduces attention, limits concentration, and lowers cognitive ability (Lufi, Tzischinsky, and Hadar 2011; Meijer, Habekothe, and Van Den Wittenboer 2000; Sadeh, Gruber, and Raviv 2003).

Since the publication of the Coleman Report in 1966, economists have contributed significantly to the body of research focused on understanding how school systems work. By approaching the development of students' human capital as a production process in which school resources, teachers, etc., are the inputs, we can utilize the rich empirical and theoretical tool kit developed to understand the production process of firms. Recent studies

---

[2] This is also true for Korean adolescents. When I restrict the sample to middle and high school students, their average sleep time decrease to 7.2 hours.

show that the daily school schedule, particularly schools' start time, is an important factor in the production process (Carrell, Maghakian, and West 2011; Edwards 2012; Heissel and Norris 2017), likely because it alleviates sleep deprivation.

Delayed school start time has also been gaining attention from the public, school districts, and even in the US Congress. At the national level, there have been several resolutions and bills that were introduced such as H. Con. Res. 176: Zzz's to A's Resolution in 2009 and H.R. 1306: Zzz's to A's Act in 2015. Recently, California SB328, which requires secondary schools to start no earlier than 8:30 AM, has been approved by the state's Senate and the Assembly Education Committee. Some school districts have already adopted delayed school start time.

However, estimating the causal effects of delayed school start time is challenging because of possible confounding factors. For example, a simple comparison of students with an early school start time and those with a later school start time may be biased by endogeneity. Unless students were randomly assigned to those two different schools, students could select into their schools based on their preferences and characteristics which may also have impacts on their sleep schedules or test scores. For instance, if a student knows that he or she is a night owl, then he or she can choose to attend a school with a later school start time. By doing so, the student can choose his or her own sleep schedule which may affect alertness in school and test scores. In this case, we cannot interpret improved test scores as being the causal result of a delayed start time.

In this paper, I use a difference-in-differences design to exploit a natural experiment in South Korea which delayed some school start times but not others. Gyeonggi Provincial Office of Education, one of the three regions within the capital area of South Korea, implemented a delayed school start time for K-12 during the 2014 academic year.[3] The policy changed school start times from 8:35, 8:20, and 8:00AM for elementary, middle, and high schools, respectively to 9:00 AM.

Korea provides a good setting in which to examine the causal impact of school start times because policy changes are implemented quickly. Rapid policy implementation does not give students and parents much time to prepare or make changes in advance. Moreover, education and student populations are fairly homogenous. Under an equalization policy, curriculum and tuition are strictly regulated by the Ministry of Education. This leads to very similar schools across regions. Furthermore, students cannot choose which school to attend because they are assigned to schools within school districts by the provincial office of education. All of these factors make it more likely that I am identifying the causal effect of changing the school start time and not some other difference between the treatment and control groups.

This study examines how a delayed school start time affects students' sleep schedules, their time use on other activities as well as academic performance, and the spillover effects on parents' time-use with the Korean Time Use Survey (KTUS) and an administrative school reporting system in a difference-in-differences framework. The KTUS is conducted every five years for representative households in South Korea. It provides a detailed time use diary for every household member over the age of nine with comprehensive information.

I find that teenagers slept an additional 33 minutes per school night on average as a result of the delayed school start time. This stems not from changes in bedtime but from

---

[3] An academic year in South Korea runs from March through the following February. However, the delayed school start time in Gyeonggi province was implemented in the middle of the 2014 academic year on September 1st.

delays in waking up. High school students gained the most sleep from this policy change (44 minutes) while elementary school students gained the least (26 minutes). I also find a spillover effect of the delayed school start time on parents. Mothers slept an additional 21 minutes and spent less time on work. Consistent with previous studies, delayed school start time in South Korea led to a statistically significant improvement of middle school students' academic performance. I find that delaying the start of school by forty minutes increased middle school math test scores by 0.034 standard deviations and reading test scores by 0.022 standard deviations. This result is interesting because the students spent less time on self-study (32 minutes) to sleep more.

Previous studies have also found a significant positive impact of delayed school start time on students' test scores (Carrell, Maghakian, and West 2011; Edwards 2012; Heissel and Norris 2017). However, none of these papers use direct measurements of students' sleep time as well as how students reallocate their time, despite arguing that changes in sleep pattern and time are the main mechanisms for students' academic improvement. Examining students' time use is important because it can provide more information about the mechanisms for academic improvement. Sleep could affect student performance through several channels: Students could sleep more. They could sleep the same amount but do so at a time more compatible with their biological clock. It is also possible that factors unrelated to sleep are involved, such as the new schedule offering better support of other academic activities. Additionally, since school start times affect parents as well as children, it is important to understand the implications of these policies for parents.

The rest of the paper proceeds as follows. Section 2 presents the background information about delayed school start time and institutional information. Section 3 describes the data sets used in this study. Section 4 presents my estimation strategies.

Section 5 presents the results including how students and their parents changed their time use and how much the policy impacted students' academic performance. Section 6 discusses robustness checks, and Section 7 concludes.

2 BACKGROUNDS

2.1 Sleep and Previous Research

Delayed timing of sleep for adolescents has long been recognized, with researchers initially attributing changing patterns in adolescence to social factors (Anders et al. 1978; Carskadon 1990; Kirmil-Grey et al. 1984). However, researchers became more focused on biological factors and found changes in circadian rhythm toward later bed and wake-up times (Andrade et al. 1993; Cardinali 2008; Carskadon, Vieira, and Acebo 1993; Crowley, Acebo, and Carskadon 2007; Wolfson and Carskadon 1998).[4][5] The sleep-wake cycle is governed by the circadian rhythm, which controls the secretion of melatonin, a sleep inducing hormone. As a child enters puberty, the circadian rhythm shifts toward later bed and wake-up times through delayed melatonin production compared to adults and younger children. The adolescent body does not produce melatonin until around 11 PM, which is 2-3 hours later than adults. This makes it difficult for adolescents to get enough sleep when they have to wake up early in the morning.

Studies find that sleep deprivation from early school start times causes not only harmful health effects and careless driving, but also negative impacts on academic performance. Edwards (2012) examined the impact of school start time on standardized tests using data on middle school students in Wake County, North Carolina. The study

4 Sleep studies show that the "delayed phase" sleep of adolescents stems more from biological factors than social factors (Crowley et al. 2007; Carskadon et al. 1993).

5 A circadian rhythm is a roughly 24-hour cycle in the physiological processes including sleep-wake timing.

exploits variation in start times within schools resulting from staggered bus schedules in response to an increased student population. It finds that a starting school one hour later increases students' math and reading test scores by 0.057 and 0.035 standard deviations, respectively. Carrell, Maghakian, and West (2011) examine the impact of the timing of a student's first class on test scores for college freshman. They utilize random assignment of cadets at the United States Air Force Academy to classes and find that the average effect of being assigned a first period course is between -0.031 and -0.076 standard deviations, and students assigned to a first period of 7AM performed 0.140 standard deviations lower, whereas a first period of 7:50 AM had an insignificant impact. Heissel and Norris (2017) exploit a time zone difference in Florida so they can study the impact of a one-hour delay in relative start times. They find that a one-hour relative later school start increases math and reading scores by 0.081 and 0.057 standard deviations for adolescents, respectively. However, none of these papers measured changes in students' sleep time or schedule as well as how students reallocate their time due to variation in school start time.

2.2 Institutional Background

South Korea consists of 17 cities and provinces. The total population is about 51 million (as of 2015), and almost half of the population lives in the capital area. GDP per capita is about 27,000 US dollars (in 2015), and the Korean people have placed high value on education. Compulsory education starts with elementary school, from 1st to 6th grade, and ends with middle school, 7th to 9th grade. However, almost all students go to high school (10th to 12th grade), and students can choose between vocational and academic high schools. An academic year starts with the spring semester in March, and the fall semester begins in September. Most Korean students enter school at six years and only in a rare situations are students held back.[6] Consequently, the Korean grade cohort is kept relatively stable.

This study exploits a natural experiment in South Korea to examine the causal effect of delayed school start time on sleep schedule and academic performance. The natural experiment is a sudden adoption of delayed school start time for K-12 in Gyeonggi province in the capital area during the 2014 academic year. For a control group, I use Seoul. Seoul is the capital city of South Korea and is surrounded by Gyeonggi province. Seoul and Gyeonggi are the two main components of the capital area in Korea. Even though these two areas cover 10.7% of South Korea, 42% of middle school students lived in Seoul or Gyeonggi during the 2014 academic year.

Nationally, schools initially resisted delaying the school start time to 9 AM. However, the Gyeonggi Provincial Office of Education made these changes a priority. Most schools in the province adopted the policy starting in the 2014 fall semester. In Seoul, the Metropolitan Office of Education did not prioritize these changes. No schools adopted the policy in 2014 and in 2015 only a few middle schools (3.7%) adopted it.

The education system in South Korea provides a good environment to examine the causal impact of the delayed school start time for two reasons: policies can be implemented in a very short time, and there exists much less variation in school education and student population compared to the United States.

In Korea, the delayed school start time was first proposed in June 2014 by students in a middle school on a bulletin board for suggestions on the Gyeonggi Provincial Office of Education website. The office reviewed the policy and gathered opinions during the summer. Then, in August, the Gyeonggi Provincial Office of Education decided to delay

[6] Students must turn six by the end of February.

school start times to 9AM for all elementary and secondary schools in Gyeonggi Province. Despite the sudden introduction, implementing ratios in the Fall semester 2014 were high: 1,155 elementary schools out of 1,195 (96.7%), 571 middle schools out of 604 (94.5%), and 302 high schools out of 451 (67.0%) delayed their start time to 9 AM starting September 1st. This rapid policy implementation would not provide enough time for students and parents to prepare or make any changes in advance.[7]

Also, South Korea has been subject to the Equalization Policy for middle schools since 1971. Under this policy, curriculum and tuition are strictly regulated by the Ministry of Education to insure very little difference across schools. Moreover, students cannot choose which school to attend because they are assigned to schools within the school districts by the office of education. Therefore, this setting makes parents unable to choose which school their child attends, which can be interpreted as near 100 percent compliance.[8]

Before the school start times were delayed, the average school start times in Gyeonggi province were about 8:35, 8:20, and 8:00 AM for elementary, middle, and high schools, respectively.[9] As illustrated in Figure 1.1, there was a morning meeting during which teachers checked attendance and made announcements.[10] Classes for elementary schools started immediately afterwards, while middle and high schools had study hall between a school start time and a first class start time. All students then attended classes with a break for lunch. The length and number of classes increase with the school level. As

7 I verify this in the robustness section using patterns of student mobility across provinces.

8 Although difficult, it is not impossible to move to a different school. If a student moves to a different school district, then the student will be reassigned to a school in the new district by the education office. As the policy was implemented in a province, students must move across provinces to avoid the policy. I examine student mobility as a robustness check and find no significant changes due to the policy.

9 This is calculated from a survey of all the primary and secondary school start times in 2014 by Gyeonggi Provincial Office of Education.

10 This figure is based on a telephone survey of representative schools. The results are consistent with the time use data.

a result, the school day ended at 2:30, 3:00 (or 3:55), and 4:10 PM for elementary, middle, and high schools, respectively.

After the delay, most of the schools in Gyeonggi Province start at 9 AM. As Figure 1.1 shows, schools did not change students' class time. Elementary schools just shifted their existing schedule back. Middle schools and high schools pushed their schedules back and also eliminated the study hall period, during which students worked by themselves.

3 DATA

The primary dataset for this paper is the Korean Time Use Survey (KTUS). This dataset contains a detailed time use diary for every household member over the age of nine for a representative sample of households in Korea. KTUS has been conducted every five years by the Korea National Statistical Office since 1999, and the first survey contained about 17,000 households with around 46,000 individuals. The first two surveys were conducted in September, whereas the third survey in 2009 was conducted in March and September for different households. The 2014 survey was conducted in July, September, and November. I use all time diaries from 1999 through 2014, except for July 2014. The July 2014 survey is omitted because it was conducted during both the final exam periods and/or summer vacation.

In addition to the time diaries, the KTUS also includes comprehensive individual and household characteristics. The diary starts at midnight and includes every activity (or activities) done (with whom) every ten minutes for 48 hours. Bed time is observed on the first night, and wake-up time is observed on the second day. Sleep duration is measured as a duration between bed time and wake-up time.[11] Since a time-use diary includes two

[11] This night's sleep duration is usually less than the total amount of sleep in a day, since the night sleep measure does not include napping and drowsing.

consecutive days, to cover all possible days in a week, there are five combinations: Sunday-Monday, Tuesday-Wednesday, Thursday-Friday, Friday-Saturday, and Saturday-Sunday. To examine how delayed school start time affects bed and wake-up times, among those five combinations, I restrict the main sample to the first three combinations to focus on school days.[12]

Figure 1.2 shows how long students engage in each activity during a typical week day. Activities for teenagers can be divided into five categories: personal activities, educational activities, leisure, household production, and the other activities. The three main activities for teenagers are personal activities (43.23%), which include sleeping and having meals; educational activities (36.89%), which consist of in-school and out-of-school activities; and leisure (12.63%). Students spend much less time on household production (0.42%) and other activities (6.83%).

Table 1.1 shows the average sleep schedules during the pre- and post-periods in Gyeonggi province and Seoul metropolitan city. There are 982 students in the pre-period and 183 students in the post-period in Gyeonggi, and 1036 students in the pre-period and 179 students in the post-period in Seoul for the time use survey. In general, older students wake up earlier and sleep later, so they sleep less, and school level differences are greater regarding bed time compared to school-level differences for wake-up time. This can be a result of delayed circadian rhythm during adolescence. The first difference between the treatment and control regions tends to be greater in the post-period than in the pre-period, especially for sleep duration and wake-up time. This results in a wake-up time that is a 0.43-hour (25.8 minutes) later and a 0.46-hour (27.6 minutes) increase in sleep duration, both of which are statistically significant at the 1% level.

---

12 The impacts of delayed school start time on sleep schedules during weekends show similar patterns as on weekdays.

The second data set is an online school information reporting system, which is collected from the Korean Education & Research Information Service (KERIS). This is used to identify test scores. Since December 2008, the Ministry of Education (MOE) has required schools to provide a variety of school level information, that is open to the public. I collected the data from 2013 to 2015 for all middle schools in Gyeonggi and Seoul.

Academic performance is reported as the mean and standard deviation of test scores for each course offered at each school by year, semester, and grade level. Test scores range from 0 to 100. The data provide the number of students tested, the average score, and the standard deviation of the score. I then normalize each school grade score by the population distribution.[13] Although schools design their own assessment metrics, the scores are comparable across schools because the equalization policy standardized curriculum and textbooks. The Ministry of Education sets the curriculum and approves textbooks for all of South Korea.[14] Also, the test scores are important for students, because those test scores are used to enter special-purpose high schools and autonomous high schools.

When examining the impact of the policy on students' academic performance, I focus on middle school students' test scores, because the implementation ratio of the policy in Gyeonggi is much higher in middle school, and the curriculum is more standardized for middle schools than high schools.[15] Moreover, there are fewer selection issues in middle schools than high schools, since there are more exceptions to the equalization policy in high schools, such as vocational high schools, special purpose high schools, and

13 When weighted by the number of students in each school, this produces estimates at the same level as student level normalized data.

14 Elementary school students even use government published textbooks.

15 For example, there are various elective subjects in high schools, whereas almost none in middle schools.

autonomous high schools.[16] In addition, for high schools more than half of the cities in Gyeonggi province are not subject to the equalization policy.

Table 1.2 shows middle school students' average test scores for the pre- and post-periods in Gyeonggi province and Seoul metropolitan city. There are 383 middle schools in Seoul and 599 middle schools in Gyeonggi. In general, math test scores show larger differences between treatment and control groups. Test scores are greater in Seoul during the pre-period; however, in the post-period, the difference becomes smaller or even the opposite for reading. This results in a 0.76-point increase in math and a 0.35-point increase in reading test scores, which are statistically significant at the 1% level.

4 ESTIMATION STRATEGY

4.1 Difference-in-Differences

To measure the causal effect of delayed school start time on sleep schedule and other time uses, I estimate the following equation:

$$Y_{isymd} = \beta_0 + \beta_1 treatment_s * post_y + \beta_2 treatment_s + \beta_3 post_y \\ + \mu X_{isy} + \gamma_y + \delta_m + \varphi_d + \varepsilon_{isymd} \tag{1.1}$$

Y_{isymd} is the dependent variable for student *i* in province *s* at year *y*, month m, and survey day d. The variable $treatment_s$ is an indicator equal to 1 if student i lives in the treated region, Gyeonggi province, and 0 otherwise. The variable $post_{\mathrm{y}}$ is an indicator equal to 1 if student i was observed since 2014 and 0 otherwise.[17] Thus, the key coefficient of interest is β_1 which measures the difference-in-differences estimate. A vector of

[16] This is important for consideration, because KTUS does not provide which type of school a student attends beyond the school level. For example, an average initial school start time for vocational high schools was 8:20AM, whereas that of education high schools was 8AM. This different treatment in delaying school start time makes it difficult to investigate the possible mechanisms for the impact of start time on test scores.

[17] Since I exclude the July survey in 2014, the remaining 2014 surveys are conducted in September or November, which occur after the treatment.

individual-specific control variables X_{isy} includes individual information such as school level, age, gender, and a dummy for living on a farm. School level dummies are included to control for school-level differences in sleep schedule, as they have different school schedules. Age and gender dummies are included to control for any age and gender differences within school level. Whether living in a farm house or not is included, because this factor would be more influenced by weather including sunrise and sunset times.[18] A year-fixed effect γ_y and a month-fixed effect δ_m are included. Since some of the surveys were also conducted in March and November, these month dummies are intended to control for seasonality, such as changes in sunrise and sunset times. A survey day-fixed effect φ_d is intended to control for any day specific differences.

4.2 Assumptions

Difference-in-differences estimation assumes that the control group is comparable, and this can be indirectly tested through the existence of parallel trends prior to the policy implementation. This is highly likely in this natural experiment, as Seoul is surround by Gyeonggi province, and they are the two main components of the capital area in Korea. In this section, I present parallel pre-trends visually through graphs, and more formal tests are presented in the robustness section.

With regard to sleep patterns, Figure 1.3 shows trends in average wake-up, bed times, and sleep duration in Gyeonggi and Seoul from 1999 to 2014 over five years. 1999, 2004, and 2009 are the pre-period, and 2014 is the post-period as the policy was implemented in September of 2014. Parallel trends in the pre-period are found in sleep schedules, thus supporting the comparability assumption. Wake-up times in the pre-period show similar trends, as differences of the two regions are around 3 to 6 minutes, whereas

[18] Rural dummy cannot be used since 2014 is the only survey that contains this information.

the difference becomes 22 minutes with an opposite sign in the post-period. Sleep duration in the pre-period show similar trends with 7 minutes of difference. However, the difference increased to 32 minutes in the post-period. On the other hand, bed time shows similar trends- even after the policy implementation. The differences in the two areas are around 10-13 minutes during the pre- and post-periods, with the exception of 3 minutes in 2009.

For test scores, Figure 1.4 shows the trends in the middle school students' test scores by semester in Gyeonggi and Seoul for reading and math. The 2013 Spring to 2014 Spring semesters are the pre-period, and the others since fall 2014 are the post-period. Parallel trends in the pre-period are also found in those subjects. Regarding reading test scores, the average test scores between the two areas are even closer until spring semester 2014, whereas there exist some differences since the policy implementation. Regarding math test scores, the differences between the two areas are more than 1 point in the pre-period, but the differences decreased after the policy implantation, as relative scores increased in Gyeonggi province.

5 RESULTS

5.1 Impacts on Sleep Schedule

Table 1.3 shows the impact of the policy on bed time in minutes. Column (1) presents the baseline estimate of the difference-in-differences, while only controlling for some covariates such as school level dummies and month and gender dummies. Considering the other parameters, we can see that middle (high) school students go to bed 40.14 (81.56) minutes later than elementary school students on average over the whole period. Moreover, age matters in terms of bed times, even with the same school level for middle and high school students. In other words, older middle (high) school students go to bed later compared to younger middle (high) school students; this finding is consistent with the scientific literature. I find that the policy does not have a significant impact on when students go to bed. This estimate is robust to the inclusion of a variety of fixed effects. Column (2) adds year effects, Column (3) includes additional age effects to control for any age-specific effects within the same school level, and Column (4) adds additional day-fixed effects. However, the effect on bed time is small and insignificant in all estimations. This null effect may reflect an already late bed time in Korea.

The impact of the policy on wake-up time is presented in Table 1.4, and it is organized like Table 1.3. According to the baseline estimate in Column (1), students could wake up 26.51 minutes later on average because of the later school start time. This statistically significant impact is robust to the inclusion of various fixed effects as in Table 1.3. Turning to the other parameters, we can see that middle (high) school students wake up 16.41 (45.19) minutes earlier than elementary school students on average. These differences are consistent with the differences in school start time by school level. I also find that, conditional on school level, student age has little impact on wake time. This makes sense because students have the same school start time in a school regardless of their age or grade.

Table 1.5 presents the impact of the policy on total amount of sleep time in a school night, organized as in Table 1.3. Since the total amount of sleep time is measured by bed time to wake-up time, Table 1.5 shows the combined results of Tables 1.3 and 1.4. The baseline results are robust to the inclusion of various control variables as in Tables 1.3 and 1.4. Delayed school start at 9 AM caused students to sleep an additional 31.24 minutes on average (Column (1)). Middle school students sleep 56.54 fewer minutes than elementary school students, and high school students sleep 126.76 fewer minutes than elementary school students. Moreover, age matters in terms of how long middle and high school students sleep, even at the same school level, as in Table 1.3.

Examining heterogenous effects by school level is necessary because the treatment itself differs by school level -25 minutes for elementary schools, 40 minutes for middle schools, and one hour for high schools. Moreover, the severity of sleep issues biologically varies with age and results from previous tables, which show patterns by school level. Table 1.6 shows the heterogenous impacts on sleep schedule at each school level, with all effects controlled for as in Column (4) in the previous tables.[19] The heterogenous effects by school level are estimated by adding interaction terms between each school-level dummy and difference-in-differences covariate instead of the previous difference-in-differences covariate in equation (1.1).

In Column (1), I find that the effects of a delayed school start time to 9AM increased with school level. Elementary school students slept 19.09 more minutes in the morning, whereas high school students slept an additional 37.25 minutes in the morning. Note that the implementing ratio in high school is only 67%, so the estimated effect is intent-to-treat for high school students. Thus, the effect of treatment-on-the-treated is 55.60 minutes for high school students. The impact for middle school students was 26.70 minutes, which falls between the elementary and high school estimates. This is reasonable because starting school at 9AM in high school delays it 60 minutes, whereas elementary school delays only 25 minutes. Column (2) verifies that the impacts on bed time are small in magnitude and statistically insignificant at all school levels. Lastly, Column (3) shows that, owing to the delayed school start time, high school students could sleep an additional 44.32 minutes, middle school students could sleep 30.48 more minutes, and elementary school students could sleep 26.32 more minutes on average.

[19] Results are robust to the inclusion of year, age, and day fixed effects as in Tables 1.3, 1.4, and 1.5.

5.2 Impacts on Other Time Use

Next, I investigate which activities students forgo in order to sleep more. Since I focus on the impacts on middle school students' academic performance, I also focus on the time use impacts for middle school students. Figure 1.5 presents changes in time use in treatment and control regions for middle school students. On average, middle school students in the treatment region slept an additional 41 minutes and spent 55 minutes less on educational activities after the policy implementation, whereas middle school students in the control region spent 10 more minutes sleeping and 6 more minutes on educational activities. To eliminate possible time trends during the time frame and different month effects, I then examine their time use on other activities with a difference-in-differences estimation that is similar to slightly modified equation (1.1) in the previous section.

Table 1.7 shows the impact of the 40-minute delayed school start time on middle school students' time use. First, I estimate the effects on six different activities in Panel A. I find that middle school students spent 29.84 more minutes sleeping, including naps, and 13.48 more minutes on non-sleep personal activities, such as having a meal, and they could do that by reducing their time spent on educational activities by 24.54 minutes and leisure by 18.77 minutes. Panel B shows which educational activities are decreased. The main difference in educational activities is decreased self-study in school by 31.79 minutes.[20] This reduction stems from the fact that students initially had a self-study period before their first class starts. Since students go to school by 9 AM and the first class starts at 9:10 or 9:20AM after the delay, schools removed the study hall period in the morning so that the first class and time to go home from school were partially delayed compared to delaying the school start time, as shown in Figure 1.1. Another interesting difference is the increased

---

20 One thing to note is that school class time does not change significantly. Even though there seems to be a 11.93-minute increase in class time, it can be a measurement error, as a class in middle school is 45 minutes and the unit of time use survey is 10 minutes.

time on non-sleep personal activities, which includes having a meal, and this is reasonable because students are more likely to have breakfast when the school start is delayed by 40 minutes and they sleep only an additional 27 minutes in the morning. Because students return home around 20-30 minutes later without changing bed time, students would have lost available afternoon or evening time. Students may have lost leisure time or out-of-school structured study time, such as private or group tutoring or online lectures. However, the estimates on these activities are not statistically significant and should be considered suggestive. Therefore, delayed school start time in Korea increases sleep time at the cost of time spent in self-study.

Moreover, since school start times affect parents as well as children, it is important to understand the implications of these policies for parents. The first spillovers of interests are changes in sleep time, because sleep patterns present the most significant change for children, followed by other activities. These spillover effects are estimated similarly to equation (1.1), but now I include more control variables such as school level, industry, and job classification for parents as well as children and marriage information. The children and marriage information needs to be controlled for, because one's time use would be affected by other family members.

Table 1.8 shows the spillover impacts of delayed school start time on parents' sleep duration. Column (1) includes both parents, whereas Columns (2) and (3) only include mothers and fathers, respectively. Column (1) shows that mothers slept 28.97 minutes less than fathers per night on average. However, having a treated child, having more than one kid, and being a single parent do not have a significant effect on parents' sleep time. On the other hand, having a treated child increased mothers' sleep time by 20.59 minutes, having more than one child decreased mothers' sleep time by 8.45 minutes (Column (2)), and being a single parent decreased fathers' sleep time by 59.33 minutes (Column (3)).

This likely reflects Korean culture, in which mothers take on most of the household production and family care.

Table 1.9 presents the spillover effects on other activities for mothers (Panel A) and fathers (Panel B). In Panel A, all of the estimates are statistically insignificant for mothers, but time spent on market work has the largest magnitudes. Market work includes doing one's main work, doing side work, taking a rest during work, doing family business without pay, etc. When I focus on the time spent doing one's main work, even though it is still statistically insignificant, the magnitude increased so that the estimate becomes -19.83, which is close to the opposite of the spillover effect on sleeping. In Panel B, the spillover effect on non-sleep personal activities is the greatest in magnitude, and it is the only single broad category that has a statistically significant impact. Fathers spent 22.33 minutes more in non-sleep personal activities on average. Personal activities are personal maintenance activities such as sleeping, having a meal, engaging in health maintenance, etc. The most plausible interpretation would be the delayed school start time increasing the time spent having a meal with family members. Even though it was unable to estimate the impact on family meals, as 2014 KTUS does not provide the detailed meal code anymore, on average having a meal was increased by 12.64 minutes, and this is significant at the 1% level. The other activities category experienced the second largest impact, among these the bulk of the time change comes from educational activities. Leisure was also decreased by 13.78 minutes, but this is statistically insignificant.

5.3 Impacts on Academic Performance

To measure the effect of later school start time on middle school students' test scores, I run a regression similar to equation (1.1) but weighted by the number of students in each school to generates student-level estimates comparable to the previous results.

School-specific characteristics are controlled for, such as whether a school is in a rural area and whether students take the test after they experienced a "free semester" in 7th grade.[21] Year-fixed effects and other factors that may affect the curriculum or tests are also controlled for, such as subject, school or school district, grade level, and the second semester in an academic year.

Table 1.10 shows the impact of delayed school start time on standardized scores. Columns (1) through (3) contain two major subjects, Columns (4) through (6) contain only reading scores, and Columns (7) through (9) contain only math scores. Columns (1), (4), and (7) present baseline estimates of the difference-in-differences without controlling for school district fixed effect and school fixed effect. In Column (1), I find that a forty-minute-later school start time increased test scores by 0.028 standard deviations on average. The impact is greater in math, with a 0.034 standard deviations increase (Column (4)), and the impact is smaller in reading, with a 0.022 standard deviations increase (Column (7)). The results are robust to the inclusion of school district or school fixed effect and all estimates are statistically significant at the 1% level.[22]

Consistent with previous studies, delayed school start time in Korea had a positive and statistically significant effects on test scores. According to Edwards (2012), a one-hour-later middle school start time increases math and reading test scores by 0.057 and 0.035 standard deviations, respectively. As stated by Heissel and Norris (2017), a one-hour relative later school start time increases math and reading scores by 0.081 and 0.057 standard deviations for adolescents, respectively. The results are also comparable to those

---

21 Some middle schools adopt a "free semester" in which7th grade students (first year in middle school) do not take midterm or final exams during their second semester. Instead, students enjoy various activities such as art, physical activities, and vocational education. This new policy was phased in from 2013 and became mandatory by the 2016 academic year. (See Chapter 3)

22 The results on log test scores are similar (Appendix Table 3).

of previous studies in several ways: Delaying school start time has greater impacts on math than on reading. Since students start to enter puberty during their middle school years, the effects are expected to be smaller for middle school students than older students, and the results in this study are smaller than those of the other study, which examined older students.[23]

6 ROBUSTNESS

6.1 Parallel Time Trend

The key assumption for a difference-in-differences strategy is the parallel time trend. In other words, the outcome of interest should follow the same trend in the absence of the treatment. Even though I cannot test this assumption directly, if the assumption holds then pre-treatment trends should be the same. With multiple periods of pre-treatment data, I can test this in a simple regression frame work. In equation (1.1), I include the common time trend variable and treated area-specific time trend instead of year and semester-fixed effects. Then I restrict the sample only to the pre-period, excluding any post-period dummies. The parallel trend assumption requires that estimates of the treated area-specific time trend should be close to zero and statistically insignificant. I run three different regressions for wake-up, bed, and sleep time with all the controls. Then, I run three different regressions for reading and math, reading only, and math only with all the controls in a similar way.

Table 1.11 presents the time trend results for three different sleep schedule measures in Panel A and for test scores in Panel B. In all three regressions in Panel A, common time trends have statistically significant estimates, whereas treated region-

[23] Please note that both the duration of the delay and the treatment effect on test score in this study is 2/3 that of in Edwards (2012). If South Korean middle schools delayed for an hour, the results might be the same.

specific time trends have much smaller estimates, which are also statistically insignificant. Panel B shows the same results. This suggests that schools that will be treated are not different from those that will not be treated.

Another method to analyze parallel time trends is to include leads in the difference-in-differences model as Autor (2003) performed. I include two leads in equation (1.1) as below:

$$Y_{istmd} = \alpha_0 + \sum_{\tau=-2}^{-1} \beta_\tau D_{s\tau} + \beta_0 D_{s0} + \gamma_s + \gamma_t \qquad (1.2)$$

$$+ \mu X_{isymd} + \delta_m + \varphi_d + \varepsilon_{isymd}$$

where τ is a relative timing with normalized to 0 at the adoption year, so a negative value indicates a lead. I then also add lags to analyze the parallel time trend in test scores as multiple post-periods are available. Parallel time trends require that leads have estimates which are close to zero and statistically insignificant. Including lags allows me to examine the post-treatment effects and their persistence.

Sleep schedule results are shown in Figures 1.6 through 1.8. Coefficient estimates of the leads and the treatment year are presented with 95% confidence intervals. In all three figures, the 95% confidence intervals of leads contain zero, which means leads are statistically insignificant at the 5% level. On the other hand, zero falls outside of the confidence interval for treatment year in wake-up time and sleep duration analysis.

Test score results are presented in Figure 1.9. Coefficient estimates of the leads, the lags, and the first treatment semester are presented with 95% confidence intervals. The confidence intervals of leads contain zero, which means leads are statistically insignificant at the 5% level, whereas the estimate of the first treatment semester has a positive value and is statistically significant at the 5% level. Estimates of lags show that the treatment effect on test scores remains relatively constant.

6.2 Check on Student Mobility

Even with parallel trends in the pre-period, the results in this paper can be subject to bias if there is endogenous student mobility. As mentioned in the institutional background section, if a student moves to a different school district, then the student will be re-assigned to a school in that new school district by a regional education office. As the policy was implemented at the province level, students must move across provinces in order not to comply with the policy. This section checks whether there was a significant change in student mobility across provinces that may bias the estimated results. I collected data from the Korean Education Statistical Service, which provides the annual number of transfer students within a province and across provinces at the administrative district level.[24] Figure 1.10 shows trends in how many middle school students transferred across provinces from 2011 to 2016 in Gyeonggi and Seoul. Trends are both before and after the policy implementation. Difference-in-differences of means is used to check whether there was a statistically significant change from the policy that is similar to equation (1.1). Since this is annual data, the post-period refers to the 2014 academic year and later. The results are shown in Table12, and there were no significant changes in student mobility from the policy.[25]

7 CONCLUSION

In this paper, I examine how a delayed school start time affects students' sleep schedules, other-time-use, other household members, and students' academic performance by exploiting a natural experiment that delayed school start times within a large province

---

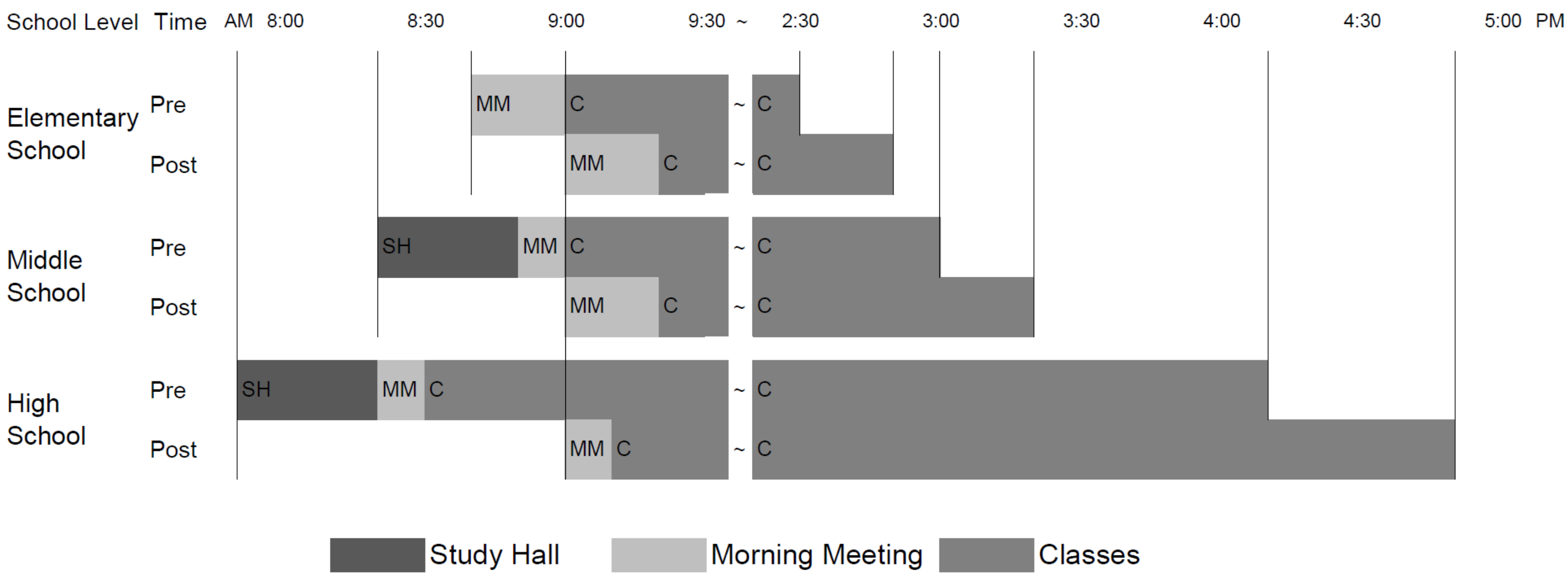
24 The data does not track which province students transferred to or came from. However, this wouldn't be a big problem, as there has long been a preference to stay within the capital area among Koreans, which results in almost half of the population living in the capital area.

25 The results are the same even when I consider since 2015 academic year as the post-period.

in South Korea. Teenagers and middle school students slept an additional 33 and 30 minutes on average per school night due to later school start times, and middle school students slept more by spending less time on educational activities in the morning. Despite this, I find that delaying school start time by 40 minutes in middle school increased math test scores by 0.034 standard deviations and reading test scores by 0.022 standard deviations. The impacts on academic performance are consistent with previous studies. Moreover, the later school start time also had spillover effects on parents. Mothers of teenager students also slept an additional 21 minutes on average per school night and spent less time on their main work.

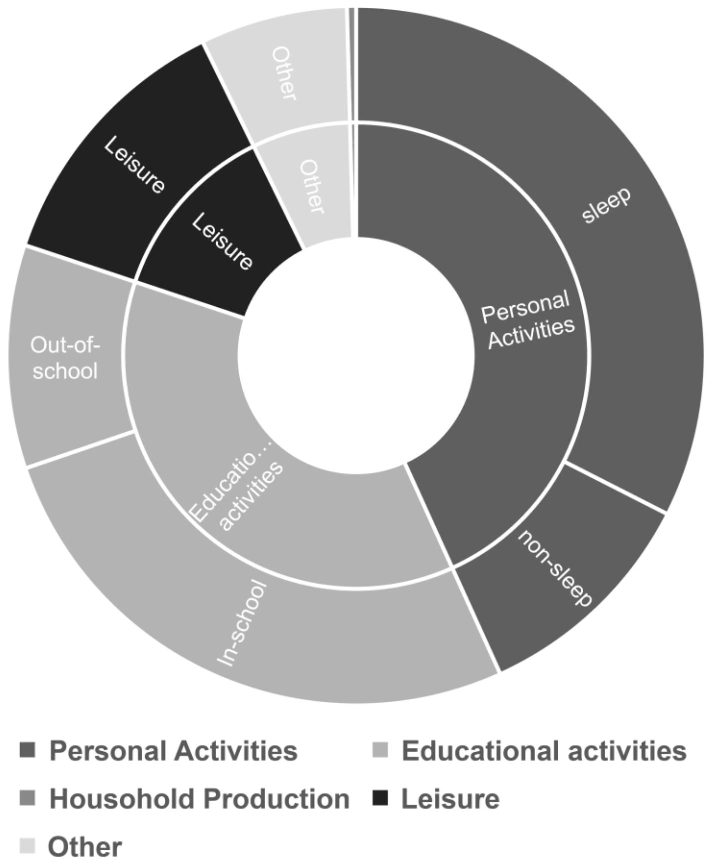
It is notable that students earn better test scores despite spending less time studying. Since previous research could not examine how students reallocate their time, it is possible that students could have used the extra time to study rather than sleep. I find, however, that students actually spent less time on self-study in school. This stems from the fact that many middle schools removed the self-study period. Students could have made this up by studying at other times during the day, but they chose to sleep more instead. Students are so sleep deprived that on the margin, sleep is more academically productive than studying.

Figure 1.1: Changes in School Schedule by School Level before and after Policy Change



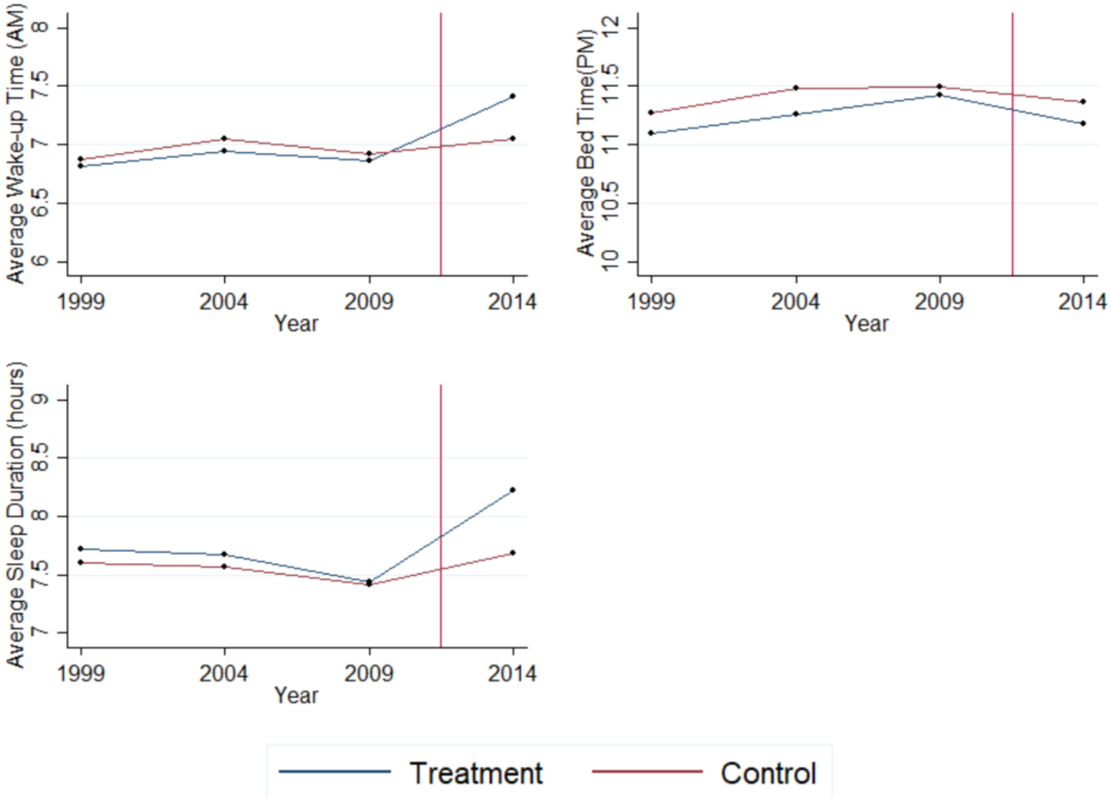
Notes: An example of changes in school schedule by school level. Source: a telephone survey of representative schools.

Figure 1.2: Classification of Time Use and Its Proportions for Teenagers in a Weekday



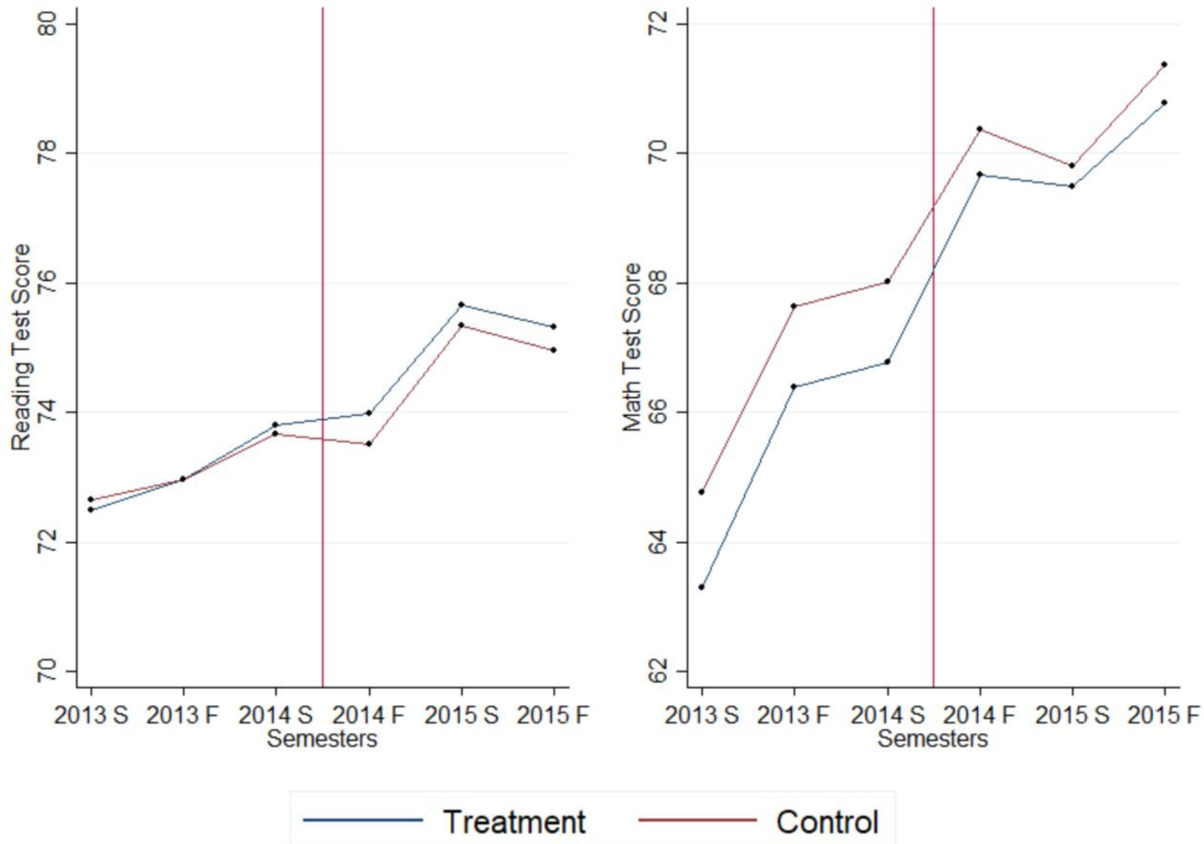
Notes: Average weekday time use for primary and secondary students over the age of nine in KTUS. Sleep includes napping and drowsing.

Figure 1.3: Trends in Sleep Schedule for Teenagers



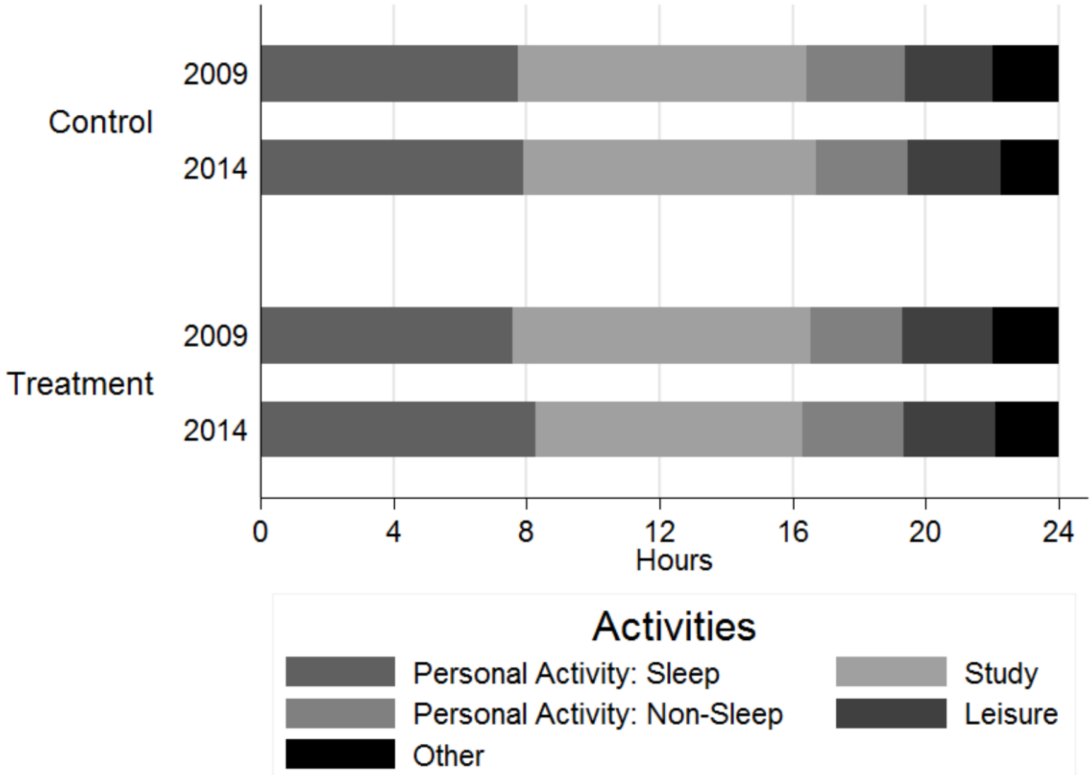
Notes: Average wake-up time, average bed time, and average sleep duration for primary and secondary school students over the age of nine are presented for treatment (Gyeonggi) and control (Seoul) from 1999 to 2014 using KTUS.

Figure 1.4: Trends in Reading and Math Test Scores in Middle School



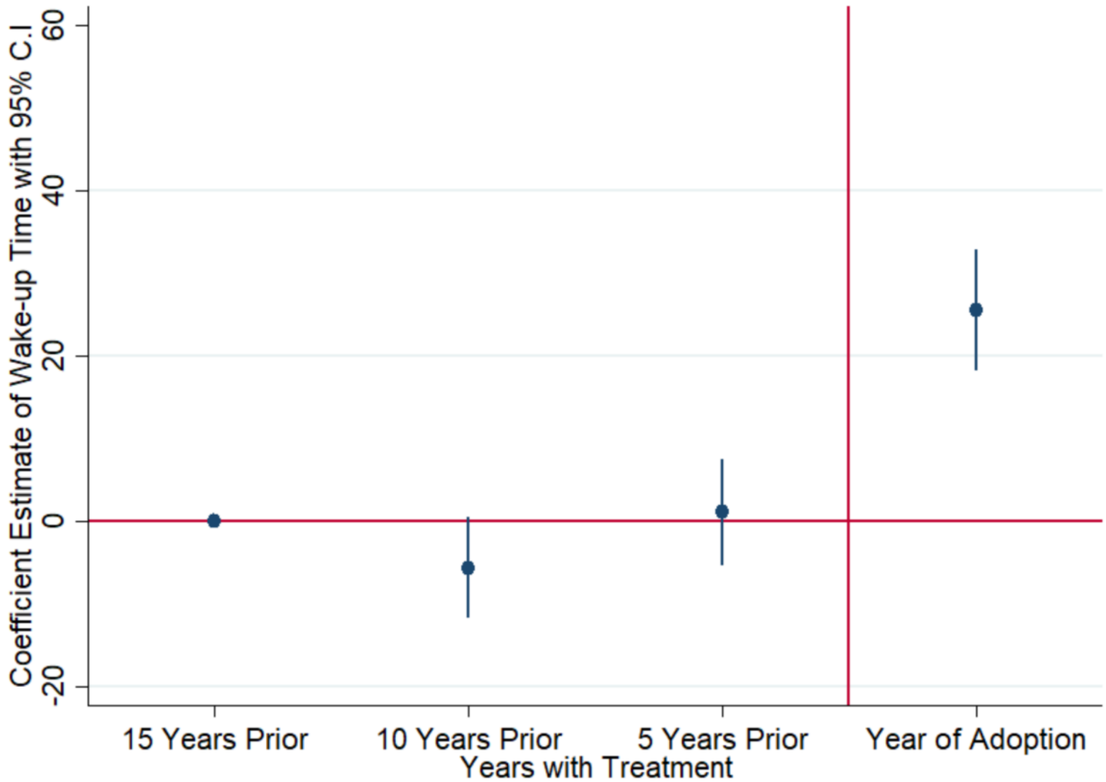
Notes: Average reading test scores and average math test scores for all middle school students in treatment (Gyeonggi) and control (Seoul) schools are presented from 2013 spring semester through 2015 fall semester. S refers spring semester and F refers fall semester.

Figure 1.5: Changes in Time Use for Middle School Students



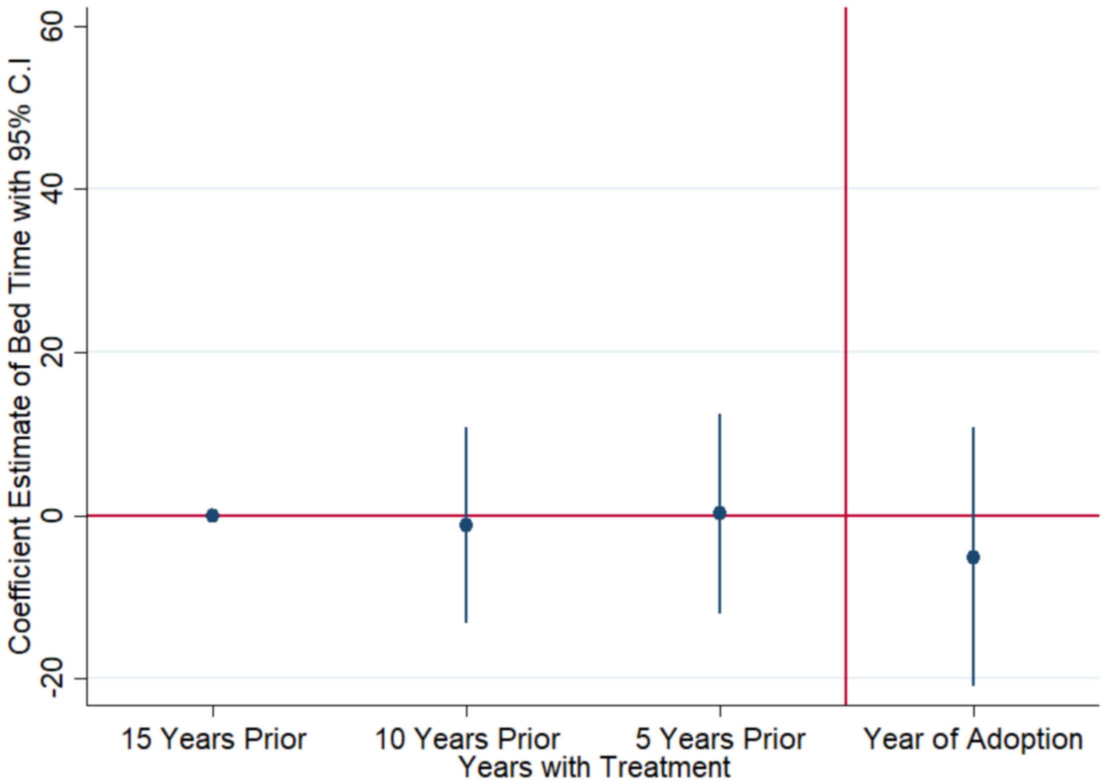
Notes: Average time use for middle school students in 2009 and 2014 for treatment and control are presented by five categories of activities. Sleep includes napping and drowsing.

Figure 1.6: Estimated Impact of Delayed School Start Time on Wake-up Time by 5 Years



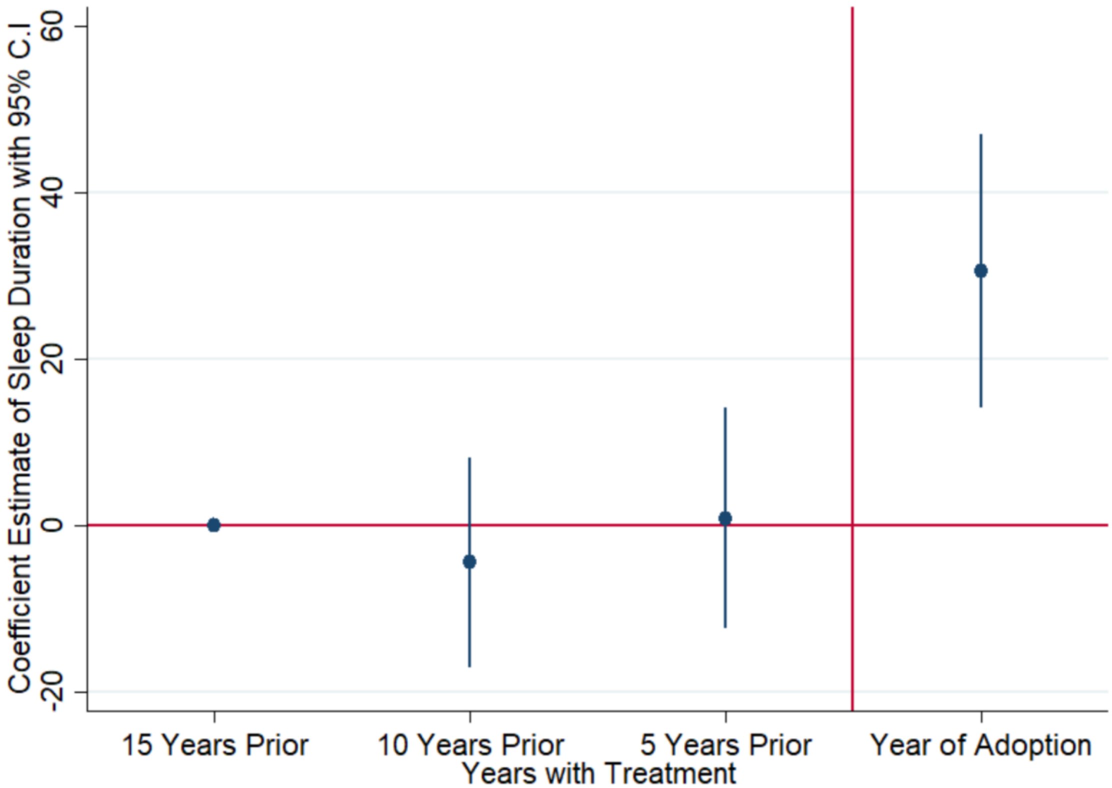
Notes: Estimated impact of the delayed school start time on wake-up time is presented with 95% confidence intervals. The dependent variable is wake-up time in minutes. Estimates are from a model that allows for effects before and during the adoption of delayed school start time. The base line effect is that of 15 years prior.

Figure 1.7: Estimated Impact of Delayed School Start Time on Bed Time by 5 Years



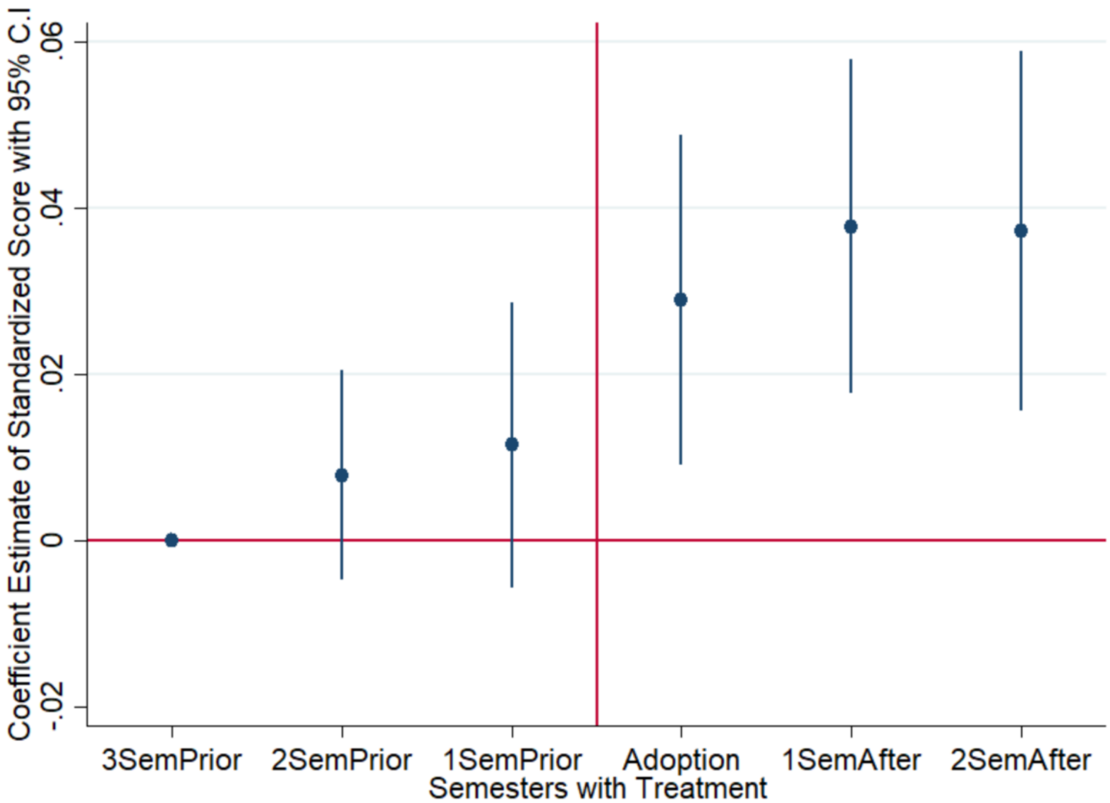
Notes: Estimated impact of the delayed school start time on bed time is presented with 95% confidence intervals. The dependent variable is bed time in minutes. Estimates are from a model that allows for effects before and during the adoption of delayed school start time. The base line effect is that of 15 years prior.

Figure 1.8: Estimated Impact of Delayed School Start Time on Sleep Duration by 5 Years



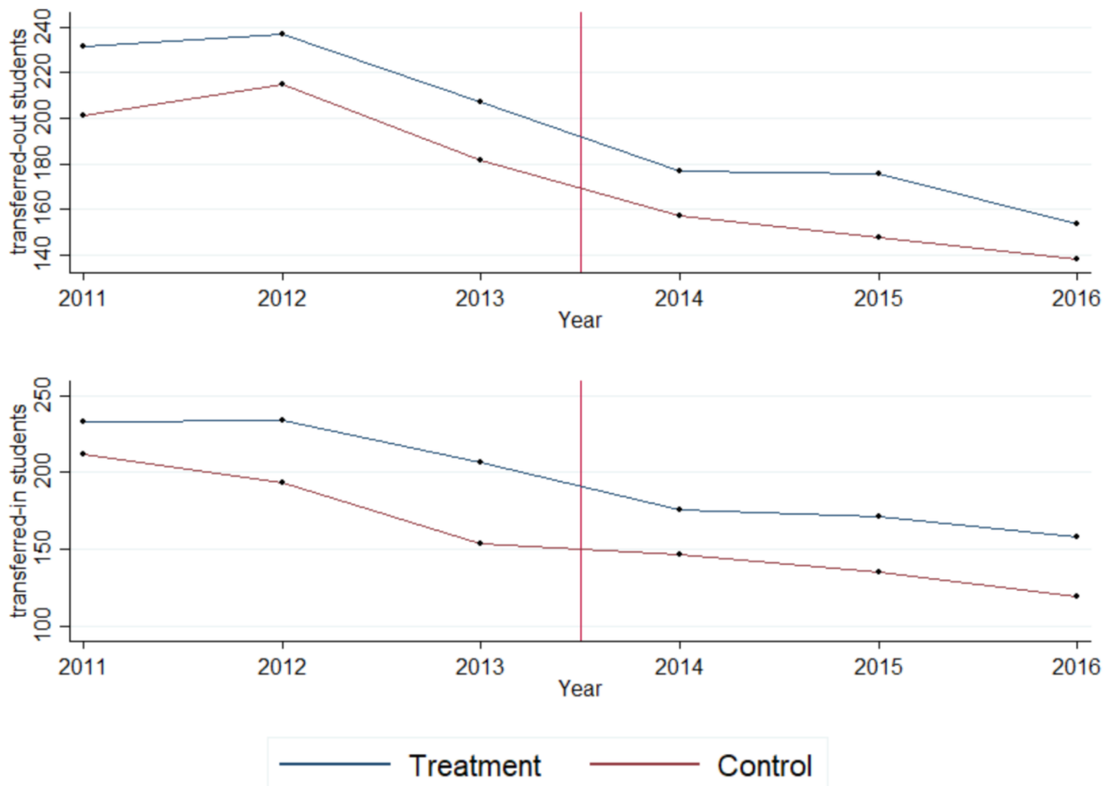
Notes: Estimated impact of the delayed school start time on sleep duration is presented with 95% confidence intervals. The dependent variable is sleep time in minutes. Estimates are from a model that allows for effects before and during the adoption of delayed school start time. The base line effect is that of 15 years prior.

Figure 1.9: Estimated Impact of Delayed School Start Time on Test Scores by Semesters



Notes: Estimated impact of the delayed school start time on test scores is presented with 95% confidence intervals. The dependent variable is standardized score. Estimates are from a model that allows for effects before, during, and after the adoption of delayed school start time. The base line effect is that of 3 semesters prior.

Figure 1.10: Trends in the Number of Transfer Middle School Students across Provinces



Notes: The average number of transferred-out and transferred-in middle school students across provinces in treatment (Gyeonggi) and control (Seoul) are presented from 2011 through 2016 by county level. Source: Korean Education Statistical Service.

Table 1.1: Average Sleep Schedules with DID without Controls

Time		Pre			Post			DID	
Region		Treatment	Control	Difference (A)	Treatment	Control	Difference (B)	DID (B-A)	t-stat
Sleep Time (hours)	All Students	7.63	7.55	0.09	8.23	7.68	0.54	0.46	3.07
	Elementary Students	8.71	8.68	0.02	9.06	8.59	0.47	0.45	2.70
	Middle School Students	7.68	7.78	-0.10	8.18	7.66	0.52	0.62	3.53
	High School Students	6.53	6.53	0.01	7.29	6.85	0.44	0.43	1.41
Bed Time (PM)	All Students	11.24	11.39	-0.15	11.18	11.36	-0.18	-0.03	0.22
	Elementary Students	10.50	10.69	-0.20	10.55	10.74	-0.19	0.00	0.01
	Middle School Students	11.22	11.26	-0.04	11.25	11.56	-0.31	-0.27	1.68
	High School Students	11.97	11.99	-0.03	11.84	11.74	0.11	0.13	0.43
Wake-Up Time (AM)	All Students	6.87	6.94	-0.07	7.41	7.05	0.36	0.43	6.87
	Elementary Students	7.21	7.38	-0.17	7.60	7.32	0.28	0.45	6.00
	Middle School Students	6.91	7.05	-0.14	7.43	7.22	0.21	0.35	3.93
	High School Students	6.50	6.52	-0.02	7.13	6.59	0.54	0.56	4.92
Observations		982	1,036		183	179			

Notes: Average sleep time in hours, average bed time in hours (PM), and average wake-up time in hours are presented for pre- and post-period in treatment (Gyeonggi) and control (Seoul). Difference-in-differences of means are presented. Robust standard errors are used.

Table 1.2: Average Test Scores with DID without Controls

Time		Pre			Post			DID	
Region		Treatment (599 Schools)	Control (383 Schools)	Difference (A)	Treatment (599 Schools)	Control (383 Schools)	Difference (B)	DID (B-A)	t-stat
Reading	All Students	73.08	73.08	0.00	74.94	74.59	0.35	0.35	2.44
	7th Grade	72.81	73.29	-0.48	74.61	74.49	0.12	0.60	2.47
	8th Grade	73.50	73.65	-0.14	74.59	74.23	0.36	0.50	2.42
	9th Grade	72.92	72.33	0.59	75.44	74.98	0.47	-0.13	0.61
Math	All Students	65.46	66.79	-1.33	69.89	70.45	-0.57	0.76	3.80
	7th Grade	66.85	68.92	-2.07	70.71	72.10	-1.38	0.69	2.65
	8th Grade	65.82	66.71	-0.89	70.52	70.58	-0.06	0.83	3.20
	9th Grade	63.75	64.85	-1.10	68.84	69.46	-0.62	0.48	1.86
Observations		10,778	6,842		9,528	6,134			

Notes: Average test scores for reading and math are presented for pre- and post-period in treatment (Gyeonggi) and control (Seoul). Difference-in-differences of means are presented. Standard errors are clustered by school.

Table 1.3: Average Treatment Casual Effect on Bed Time

	(1)	(2)	(3)	(4)
VARIABLES	Bed Time			
Mean	11:18 PM			
Treated*Post	-4.729	-4.083	-5.156	-5.667
	(7.455)	(7.457)	(7.370)	(7.371)
Middle School	40.136***	39.976***	24.192***	24.858***
	(2.538)	(2.517)	(5.585)	(5.557)
High School	81.561***	81.834***	33.939***	35.257***
	(3.212)	(3.201)	(7.917)	(7.889)
Year Effects		✓	✓	✓
Age Effects			✓	✓
Day Effects				✓
Observations	2,380	2,380	2,380	2,380
R-squared	0.234	0.242	0.283	0.288

Notes: Robust standard errors are presented in parentheses. Significance level does not change when using clustered standard errors. The dependent variable is bed time in minutes. Month, gender, and farm effects are controlled for in all estimation. *** 1% significant; ** 5% significant; * 10% significant.

Table 1.4: Average Treatment Casual Effect on Wake-up Time

	(1)	(2)	(3)	(4)
VARIABLES	Wake-Up Time			
Mean	6:57 AM			
Treated*Post	26.512***	26.805***	26.759***	26.993***
	(3.305)	(3.306)	(3.321)	(3.330)
Middle School	-16.407***	-16.384***	-17.072***	-17.044***
	(1.391)	(1.392)	(2.845)	(2.847)
High School	-45.194***	-44.996***	-43.413***	-43.360***
	(1.455)	(1.450)	(4.559)	(4.571)
Year Effects		✓	✓	✓
Age Effects			✓	✓
Day Effects				✓
Observations	2,380	2,380	2,380	2,380
R-squared	0.328	0.334	0.336	0.337

Notes: Robust standard errors are presented in parentheses. Significance level does not change when using clustered standard errors. The dependent variable is wake-up time in minutes. Month, gender, and farm effects are controlled for in all estimation. *** 1% significant; ** 5% significant; * 10% significant.

Table 1.5: Average Treatment Casual Effect on Sleep Duration

	(1)	(2)	(3)	(4)
VARIABLES	Sleep Duration			
Mean	7 hours 39 mins			
Treated*Post	31.241***	30.888***	31.915***	32.660***
	(7.650)	(7.652)	(7.543)	(7.549)
Middle School	-56.543***	-56.360***	-41.263***	-41.902***
	(2.631)	(2.621)	(5.759)	(5.716)
High School	-126.755***	-126.831***	-77.352***	-78.617***
	(3.367)	(3.362)	(8.440)	(8.414)
Year Effects		✓	✓	✓
Age Effects			✓	✓
Day Effects				✓
Observations	2,380	2,380	2,380	2,380
R-squared	0.406	0.408	0.438	0.443

Notes: Robust standard errors are presented in parentheses. Significance level does not change when using clustered standard errors. The dependent variable is sleep time in minutes. Month, gender, and farm effects are controlled for in all estimation. *** 1% significant; ** 5% significant; * 10% significant.

Table 1.6: School Level Specific Casual Effects on Sleep Schedule

	(1)	(2)	(3)
VARIABLES	Wake-Up time	Bed Time	Sleep Duration
Mean	6:57 AM	11:18 PM	7 hours 39 mins
Treated*Post*Elementary	19.090***	-7.226	26.316***
	(3.921)	(8.954)	(9.196)
Treated*Post*Middle	26.696***	-3.779	30.475***
	(4.251)	(8.438)	(8.968)
Treated*Post*High	37.249***	-7.067	44.316***
	(5.142)	(13.866)	(13.457)
Middle School	-17.921***	24.507***	-42.428***
	(2.935)	(5.630)	(5.788)
High School	-45.084***	35.208***	-80.292***
	(4.607)	(8.032)	(8.535)
Observations	2,380	2,380	2,380
R-squared	0.339	0.288	0.443

Notes: Robust standard errors are presented in parentheses. Significance level does not change when using clustered standard errors. The dependent variables are listed above and measured in minutes. Month, gender, farm, age, year, and day effects are controlled for in all estimations. Estimates are robust to the exclusion of these fixed effects. *** 1% significant; ** 5% significant; * 10% significant.

Table 1.7: The Effect of Later School Starting Time on Time Use (Middle School Students)

Panel A: The Effect of Later School Starting Time on Activities for Middle School Students

Personal Activities: Sleep	Educational Activities	Leisure	Personal Activities: Non-Sleep	Household Production	Other Activities
29.837***	-24.539	-18.766	13.479**	4.737*	-4.748
(8.510)	(16.605)	(13.232)	(6.123)	(2.848)	(6.886)

Panel B: The Effect of Later School Starting Time on Educational Activities for Middle School Students

In-School			Out-Of-School		
Class	Self-Study	Other	Structured	Self-Study	Other
11.932	-31.789***	5.173	-6.597	-1.539	-1.719*
(8.727)	(5.059)	(6.759)	(11.214)	(11.006)	(1.036)

Notes: Robust standard errors are presented in parentheses. Significance level does not change when using clustered standard errors. The dependent variables are the number of minutes spent doing each activity. Month, gender, farm, age, year, and day effects are controlled for in all estimations. *** 1% significant; ** 5% significant; * 10% significant.

Table 1.8: The Spillover Effect of Later School Starting Time on Parents' Sleep

	(1) Both	(2) Mother	(3) Father
VARIABLES		Sleep Duration	
Mean	7hr 14min	7hr 5min	7hr 24min
Treated*Post	10.640	20.591**	-3.232
	(7.896)	(10.015)	(12.325)
More Than 1 Kid	-0.535	-8.446**	6.324
	(3.113)	(3.952)	(4.887)
Single Parent	-12.557	12.160	-59.331***
	(18.729)	(25.159)	(22.963)
Female	-28.966***		
	(3.612)		
Observations	2,849	1,490	1,359
R-squared	0.064	0.068	0.080

Notes: Robust standard errors are presented in parentheses. The dependent variable is sleep time in minutes. Month, gender, single parent, having more than one kids, having preschooler, age, education, job classification, industry, year, and day effects are controlled for in all estimations. *** 1% significant; ** 5% significant; * 10% significant.

Table 1.9: The Spillover Effect of Later School Starting Time on Parents' Activities

Panel A: The Spillover Effect of Later School Starting Time on Activities for Mother

Market Work	Non-Sleep Personal Activity	Household Production	Leisure	Other Activities
-12.430	9.284	-6.881	-6.521	-4.044
(18.370)	(8.497)	(14.860)	(15.824)	(14.917)

Panel B: The Spillover Effect of Later School Starting Time on Activities for Father

Market Work	Non-Sleep Personal Activity	Household Production	Leisure	Other Activities
4.725	22.331**	5.442	-13.782	-15.484
(23.073)	(8.899)	(5.431)	(16.904)	(14.774)

Notes: Robust standard errors are presented in parentheses. The dependent variables are in the number of minutes spent in each activity for mothers (Panel A) and for fathers (Panel B). Month, gender, single parent, having more than one kids, having preschooler, age, education, job classification, industry, year, and day effects are controlled for in all estimations. *** 1% significant; ** 5% significant; * 10% significant.

Table 1.10: The Effects of Later School Starting Time on Middle School Standardized Test Scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Reading and Math			Reading			Math	
VARIABLES					Standardized Average Test Score				
Treated*Post	0.028***	0.028***	0.027***	0.022***	0.022***	0.023***	0.034***	0.034***	0.032***
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)
8th Grade	0.001	0.001	0.003	0.001	0.001	0.003	0.001	0.000	0.002
	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
9th Grade	0.003	0.003	0.006	0.005	0.005	0.009	0.000	0.000	0.003
	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.005)
School District Effects		✓			✓			✓	
School Effects			✓			✓			✓
Observations	33,282	33,282	33,282	16,640	16,640	16,640	16,642	16,642	16,642
R-squared	0.003	0.044	0.413	0.001	0.030	0.479	0.011	0.075	0.551

Notes: Standard errors, clustered by school, are presented in parentheses. The dependent variable is standardized average test scores. Columns (1)-(3) include reading and math test scores along with a subject indicator. Columns (4)-(6) include only reading test scores and Column (7)-(9) include only math test scores. Year, semester, “free semester”, and rural effects are controlled for in all estimations. *** 1% significant; ** 5% significant; * 10% significant.

Table 1.11: Test for Parallel Time Trend during Pre-Period

Panel A: Test for Parallel Time Trend for Teenagers' Sleep Schedules during Pre-Period

VARIABLES	(1) Wake-Up Time	(2) Bed Time	(3) Sleep Duration
Common Time Trend	3.321***	9.020***	-5.700**
	(1.239)	(2.253)	(2.449)
Gyeonggi Specific Time Trend	0.239	1.142	-0.903
	(1.630)	(3.084)	(3.349)

Panel B: Test for Parallel Time Trend for Middle School Test Scores during Pre-Period

VARIABLES	(1) Average Scores for Math and Reading	(2) Average Score for Reading	(3) Average Score for Math
Common Time Trend	1.087***	0.500***	1.674***
	(0.066)	(0.071)	(0.099)
Gyeonggi Specific Time Trend	0.124	0.147	0.099
	(0.089)	(0.095)	(0.131)

Notes: In Panel A, robust standard errors are presented in parentheses due to small number of clustering (region by school level). Significance level does not change when using clustered standard errors. The dependent variables are 3 different sleep schedule measures in minutes. In Panel B, standard errors, clustered by school, are presented in parentheses. The dependent variables are school, grade, semester, course level average test scores for math and reading, reading only, and math only. *** 1% significant; ** 5% significant; * 10% significant.

Table 1.12: Test for Middle School Student Mobility Across Provinces

VARIABLES	(1) The Average Number of Students Transferred-In Across Provinces	(2)	(3) The Average Number of Students Transferred-Out Across Provinces	(4)
Post	2014 and Later	2015 and Later	2014 and Later	2015 and Later
Treated*Post	-3.742	1.008	-5.009	-2.996
	(11.637)	(10.700)	(9.781)	(9.173)
Observations	336	336	336	336
R-squared	0.068	0.068	0.058	0.058

Notes: Standard errors, clustered by county, are presented in parentheses. The dependent variables are the number of students transferred-in across province for column (1) and (2), the number of transferred-out across province for column (3) and (4). Column (1) and (3) define post to equal 1 in the 2014 academic year and later, column (2) and (4) set post equal to 1 in the 2015 academic year and later. *** 1% significant; ** 5% significant; * 10% significant.

Chapter 2: Mixing Good and Bad Apples: Internal and External Spillover Effects of Autonomous High Schools[26]

1 INTRODUCTION

Students spend a significant portion of their time together in school, where they learn, play and eat together. In South Korea, the country of our study, academic high-school students spend on average at least 6.8 hours per weekday in school.[27] That is, they spend more than a third of their non-sleeping time in school. It is unlikely they do not get influenced by their schoolmates.

Educational policy makers, teachers, and parents all are eager to find out how students affect one another, particularly in terms of their academic performance. Thus, there have been many studies on peer effects in academic performance. Although results from studies on higher education peer effect are mixed, most of secondary education peer effect studies find at least modest effects.[28] These different peer effects depend largely on the definition of peer groups. Carrell, Fullerton, and West (2009) find that peer effects estimates can differ greatly depending on the set of peers.[29] For example, roommates and dorm floors in higher education comprise a limited portion of an individual's peer group. On the other hand, most of the studies on secondary education peer effects have focused on peer effects within a classroom or a grade. However, students peer networks can be

---

26 This is a co-work with Jungmin Lee from Seoul National University, South Korea, and IZA, Germany.

27 They have 50 minutes of 7 classes and one-hour lunch time on weekdays.

28 Peer effect studies in secondary education exploit variation across classrooms or cohorts within a school. [Betts and Zau 2004; Burke and Sass 2013; Carrell, Malmstrom, and West 2008; Hanushek et al. 2003; Hoxby and Weingarth 2005; Vigdor and Nechyba 2007]. Peer effect studies in higher education exploit situations in which individuals are randomly assigned to peer groups such as roommates or dorm floors. [Foster 2006; Sacerdote 2001; Stinebrickner and Stinebrickner 2006; Zimmerman 2003]

29 Burke and Sass (2013) also shows that peer effects are greater when defined in a classroom level than defined in a grade-level.

much larger; students could include peers in other grades, or even the schools in their network.

In this paper, we study the existence of spillover effects, that includes peer effects, between grades in high school. We attempt to answer this question by exploiting a natural experiment in Seoul, South Korea—the introduction of autonomous high schools. The country is well-known to education researchers for its Equalization Policy. This policy has three primary components: First, curricula and tuition are strictly regulated by the government, regardless of whether schools are private or public. Second, public school teachers are regularly rotated across schools to ensure an even distribution of teaching quality. Lastly and more importantly, students are assigned to schools within school districts according to a formula, which is not available to the public. The high school Equalization Policy started in Seoul and Busan in 1974, and many other parts of the country are currently subject to it.

Starting in 2010, the government introduced the concept of autonomous high schools by designating 27 existing private schools in Seoul as such. Unlike other schools, autonomous schools select their own students from anywhere in Seoul, and are not required to adhere to the equalization policy. This autonomous high school policy is part of a broader attempt to promote school choice and competition for students.

The introduction of these autonomous schools creates a natural experiment, which allows us to assess how changes in the composition of a school at the grade-level affects the performance of other students. Prior to 2010, autonomous schools were subject to the Equalization Policy, which guaranteed their composition as similar to that of other schools. However, starting in 2010, with their new designation, these schools began to recruit strong students from non-autonomous schools. Existing students prior to the designation would remain enrolled under the equalization policy. In the first year of the program, autonomous high school students were better in their first year than the incumbent students in their second or third year. This results in “cream skimming”, where in remaining first year students from the non-autonomous schools become thus weaker than students in their second or third years.

We exploit this exogenous variation in peers at the school level, and estimate its impact in the distribution of high-achieving students on incumbent students. The findings show the influx of high-achieving students in autonomous schools does not significantly affect its incumbent students. In the nearest non-autonomous schools, the influx of low achieving students significantly decreases the academic achievement of incumbent students, while they become more heterogeneous, likely due to the poor performance of lower-achieving students.

The remainder of the paper proceeds as follows: first, in Section 2, we generally explain the education system in Seoul and the background for introducing autonomous high schools. Section 3 introduces the data and empirical models that we use. In Section 4 we present and discuss the results, and Section 5 concludes.

2 INSTITUTIONAL BACKGROUND

The equalization Policy is an assignment system through lottery in South Korea for the balanced student distribution in high schools based on the abilities of students. In other words, students cannot choose to attend a specific high school; rather, they are assigned to one in their school district by random lottery. For example, if a student passes a certain threshold score on a standardized test and chooses not to attend a vocational high school, he or she will be subject to the equalization policy and randomly assigned to a high school, which can be either private or public in the student’s school district. Once the student is assigned to a high school, he or she can then transfer to a different high school in a new school district.

Both public and private high schools are under this policy so that every high school, along with its respective curriculum and tuition, is regulated and set by the government. Therefore, through balanced student distribution and the aforementioned regulations, the equalization policy can render schools equal, which legitimizes the name it bears.

This policy was initiated to reduce excessive educational costs of shadow education. As more students graduate high school and attend college in Korea, parents want their children to attend better high schools so they will be more likely to attend a prestigious university. Thus, it was too competitive to be admitted to better high schools when each high school chose its own students. To reduce those burdens, the policy was introduced in the 1970s in some provinces, including Seoul. Due to initial success, it was adapted nearly nationwide in many school districts, with the exception of some in distant rural areas.

However, the equalization policy has a major limitation as well, as there still exist demands for better quality high school education in Korea. In other words, increasing competitiveness in the admittance to prestigious colleges incite parents to place their children where there are better educational opportunities such as special-purpose high schools.

Although the demands for special-purpose high schools had increased, there were not enough of such schools in Korea, which resulted in more severe competition among middle school students. To alleviate this excess in demand for better education, a new form of high school was introduced from 2010 which promoted school choice and competition among schools,. This type of school is called an autonomous high school because they have more autonomy in terms of their choices in curriculum and teachers, compared to traditional high schools under the equalization policy. Students who meet the minimum criteria can apply to an autonomous high school.

3 DATA AND EMPIRICAL STRATEGY

3.1 Data

Our data are collected from the Korean Education & Research Information Service (KERIS), an online school information reporting system. Since December 2008, the Ministry of Education (MOE) required schools to report a variety of information through the system, such as student outcomes (course-level performance for each grade, dropout, transfers, and so on), expenditures per pupil and weekly teaching hours per teacher. The collected information is available to the public. We collected data from 2009 to 2012.

The sample was limited to high schools in the capital city of Seoul, which had implemented the equalization policy since 1974. More than half of all autonomous high schools in Korea are in Seoul. With the exception of vocational and special-purpose high schools, there are a total 197 high schools in Seoul. Among them, 13 high schools were selected to become autonomous in 2010, and 12 additional schools followed suit in 2011.[30]

To examine the spillover effects of autonomous schools and beyond, we selected non-autonomous schools wherein a spillover effect would likely occur. Specifically, for each autonomous school, a non-autonomous school closest to it was selected. We also took into account the sex of the students in the school. For a single-sex autonomous school, we selected the closest non-autonomous school with the same sex type or coed non-autonomous school. If a non-autonomous school was matched with multiple autonomous

[30] We exclude two autonomous schools out of 27 since student composition of those two school were not similar to other academic high schools before the transition. One school was not a general academic high school and had been exempt from the equalization policy even before the transition. The other school was a vocational high school before the transition.

schools, we selected the closest autonomous school. There were three autonomous schools, for which we could not find any matched school that satisfied the criteria.

The student outcome of our main interest is academic performance. The mean and standard deviation of test scores for all courses offered at each school were collected. Although different schools offer different curriculum, they are comparable across schools to an extent due to the equalization policy. As the test scores are not standardized, we will control for course-school fixed effects in all the regression analyses below. The course test scores are important for students, as individual students' rankings within schools are calculated based on those scores and are critical for college admission.

Table 2.1 presents summary statistics for the transition to autonomous (TTA) schools and matched comparison schools before and after the transition. Nearly all characteristics are the same between two types of schools before the transition, only except for school budget. This is simply a reflection of the fact that all TTA schools are private schools, which have higher school budget per pupil. This can be also verified in Table 2.2, which presents pre-transition differences between TTA and matched comparison schools. It is important to note that incumbent students in TTA schools are still the subject to the equalization policy. For example, these students continue to pay the same amount of tuition as before, though newly selected students pay much higher tuition.

3.2 Spillover Effects in TTA Schools

We want to identify two types of spillover arising from the introduction of autonomous schools. The first spillover occurs within TTA schools through the influx of higher-achieving students to the schools. As explained above, there are existing students who had entered the school before it became autonomous, and they should be mixed with incoming selected students. They are in different grades but there might be some external effects; they share the same facilities and teachers, and may meet during extracurricular activities. There will be a further discussion on potential mechanisms for cross-grade spillovers in the study. To capture this spillover effect, we estimate the following equation with the sample of only TTA schools:

$$\ln(S_{kgit}) = \beta \cdot TTA_{git} + \delta \cdot PEER_{git} + X_{kgit}\gamma + \alpha_g + \alpha_{ki} + \alpha_t + \epsilon_{kgit} \qquad (2.1)$$

where the dependent variable is the natural logarithm (ln) of mean score (or standard deviation) of course k in grade g, school i, and year t. The unit of observation is course-school-semester-year. A variable of our key interests in this estimation is the indicator TTA_{git}, which takes the value of one for the courses offered to g[th] grade students in school i and year t, who entered the school after its transition and the value of zero for those who had entered the school prior to the transition. The coefficient β captures the gap in mean score (or standard deviation) between those who entered a TTA school after the transition and those who had entered the same school prior to the transition. The existence of the cream-skimming effect must be verified first. If better students are selected into autonomous schools as expected, then β would be positive.

Another variable of our key interests, $PEER_{git}$, is the dummy variable that indicates whether students in grade g (= 10, 11 or 12), school i, and year t are mixed with newly selected students. The sign of the spillover effect δ, is an empirical question. In fact, there might be many different channels for the cross-grade spillover effect. The presence of higher-achieving students may improve the overall school environment (such as less trouble between grades or positive motivation from new cohorts) and benefit other students in the same school. The positive effect may occur directly through the interaction of students' or indirectly by teachers, as they can focus more on teaching due to improved school environment. However, there might be a negative spillover effect. Schools and teachers might reallocate their resources and efforts toward new students who are relatively better than their seniors. Also, extensive heterogeneity among students could reduce the productivity of teachers.

Vector X_{kgit} includes control variables, such as average expenditures per pupil, teaching hours per teacher, and the number of credits per course. We control for three sets of fixed effects: grade-specific fixed effect (α_g), school-course fixed effect (α_{ki}), and year fixed effect (α_t). The school-course fixed effect is controlled for to address the possibility that schools have different grading policies for various courses. For example, math teachers in one school might be stricter than those in another school, thus the mean score of the former school is lower in the absolute term than that of the latter.

Figure 2.1 shows which groups are "treated" and controlled for a TTA school of the transition year *t*. The number in a cell indicates grade (= 10, 11 or 12). The control group is indicated in gray, and consists of test scores at (t-1) or earlier of those who entered the school before year *t*. There are two treatment groups: (1) the cream-skimming effect group, identified in red (test scores at *t* or later of those who entered the school at *t* or later); and (2) the cross-grade spillover effect group, identified in blue (test scores at *t* or later of those who entered the school at (*t*-1) or (*t*-2)).

3.3 Spillover Effects in Matched Schools

The second type of the spillover effect occurs between schools, from autonomous schools to non-autonomous schools, in particular, those schools that are geographically close to autonomous schools. There may be two sub-types of the between-school spillover effect. First, as autonomous schools admit better students, non-autonomous schools in the neighborhood are left behind with lower-achieving students ("left-behind" students). This direct spillover effect is simply the mirror image of the cream-skimming effect in the previous subsection.

Second, as the average quality of incoming students in neighboring schools gets worse, incumbent students in those schools could be affected. This spillover effect across grades can be either positive or negative; as the quality of new students gets worse, teachers might reallocate their efforts in favor of incumbent students who should be relatively better. Alternatively, lower-achieving new students might deteriorate overall school environment and decrease teachers' productivity, affecting incumbent students negatively.

We estimate the two types of the between-school spillover effect simultaneously by estimating the following equation.

$$\ln(S_{kgi't}) = \beta \cdot LeftBehind_{gi't} + \delta \cdot PEER_{gi't} + X_{kgi't}\gamma \qquad (2.2)$$

$$+\alpha_g + \alpha_{ki'} + \alpha_t + \epsilon_{kgi't}$$

where subscript i' indicates that the school is the non-autonomous in closet proximity to autonomous school i. The matched non-TTA school is "treated" by school i's transition, which means more lower-achieving students would enter the matched non-TTA school because autonomous schools cream skim for higher-achieving students. This cross-school spillover effect is captured as $LeftBehind_{gi't}$ which equals TTA_{git} in the matched TTA school. If there is left-behind effect as expected, then β would be negative. $PEER_{gi't}$ is the indicator of whether students in grade g, non-TTA school i' and year t are mixed with lower- achieving newly admitted students. As explained above, the sign of the effect, δ_n, is ambiguous in theory. Figure 2.1 also demonstrates which groups are ″treated" and controlled for a non-TTA school that matched with a TTA school of the transition year t.

4 RESULTS

4.1 Average Effects

Table 2.3 shows average cream-skimming effects and cross-grade spillover effects in autonomous schools. Columns (1) and (3) present the baseline estimates for the natural log of mean test scores, and standard deviation, respectively while only controlling for fixed effects such as grade-specific fixed effects, school-course fixed effects, year and semester fixed effects. We discover the existence of the cream skimming effect through the transition to the autonomous high schools, but there was no significant cross-grade spillover effect within TTA schools. These estimates are robust to the inclusion of school characteristic control variables. According to the baseline estimates in Columns (1) and (3), there were 16.92% increase in test scores and 23.99% decrease in standard deviations for those who were admitted after the transition and those effects were significant at the 1% level. This confirms that the demands for better quality high school education makes the transition schools more appealing for better-quality students to enter. These influxes of higher-quality students did not have significant impact on test scores for the older incumbent students who had entered the school prior to its autonomous transition and were mixed with incoming selected students. According to the baseline estimates in Columns (1) and (3), they had 0.99% decrease in test scores and 1.59% decrease in standard deviations, but both were not statistically significant.

The average spillover effects in matched comparison schools are presented in Table 2.4, and organized like Table 2.3. According to the baseline estimates in Columns (1) and (3), students who entered the matched comparison school after the transition of a neighboring TTA school had 3.18% lower test scores, with 3.73% increase in standard deviations compared to those who took the tests before those new left-behind students entered the schools, and they are statistically significant. These estimates are robust to the inclusion of school characteristic control variables. Thus, this would make sense, as once TTA schools cream skim for better quality students, the left-behind students consist of less better- achieving students, compared to when all schools were subject to the equalization policy. However, since there are much less TTA schools than the remaining traditional schools, the effect would have much greater magnitudes in TTA schools than in neighboring comparison schools.

Then, decrease in the average quality of new students in matched schools could affect incumbent students in those schools. These cross-grade spillover effects are presented in the row of peer; 1.33% decrease in test scores and 2.49% increase in standard deviations were observed, but the impact on average test scores is not significant. However, once we included more control variables, the impact was 1.65% decrease in test scores, which is significant at the 10% level, thus suggesting there may be negative cross-grade spillover effects.

4.2 Heterogeneous Effects

How successfully new autonomous schools attract better-quality students could differ because students could choose whether to apply for a specific TTA school or not. If heterogeneous cream-skimming effects by school level are present, so are heterogeneous cross-grade spillover effects within TTA schools and across schools. Therefore, the existence of any heterogeneous effects is our next point of interest.

We first need to estimate the heterogeneous cream-skimming effects from the introduction of autonomous schools since any heterogeneous effects will stem from heterogeneous cream-skimming effects. We estimate the performance gap for each school by students who entered before the transition and those after, similarly to equation (2.1).

Since we wish to estimate the performance gap that results from selection into autonomous schools, we limit the sample to those courses offered to 10^{th} grade (first year high school) students in their first semester, to capture the selection effect *per se* while minimizing any subsequent effects arising after admittance, such as peer effects within a grade or school resource effects. The sample includes only TTA schools.

The average cream-skimming effects in TTA schools are presented in Table 2.5. New selected students received 8.24% higher test scores, and standard deviation decreased by 19.76% compared to those who entered the schools through the equalization process prior to the transition, and the estimates are robust to inclusion of more covariates. When the focus is only on students' first semester test scores, though the effects are still statistically significant at the 1% level, the effects are less than those in Table 2.3 where we use all the test scores. That may support the positive peer effects within a grade and/or school resource reallocation toward newly selected students. An estimation on the heterogeneous cream-skimming effects in each TTA school is then considered by allowing estimates to differ by each school. Panel A in Figure 2.2 shows how the effects differ by school level. The order is sorted by the extent of the effect, and asterisks indicate the significance level as in Tables. More than a half TTA schools were successful in attracting better incoming students, but still there exist large differences in magnitudes. For example, 4 schools experienced more than 15% increase in test scores for new incoming students, which is almost twice of the average effect in Table 2.5.

Furthermore, such heterogeneous cream skimming may have caused heterogeneous left-behind effects in matched comparison schools, which were estimated in similar fashion to the analysis for cream-skimming effects. Table 2.6 presents the average left-behind effects. Left-behind incoming students received 3.09% lower test scores, and standard deviation increased by 6.84% compared to those who entered the comparison schools prior to the transition of matched TTA schools, and the estimates are robust to inclusion of more covariates. We then estimate heterogeneous left-behind effects in each matched comparison school by allowing estimates to differ by school level. Panel B in Figure 2.2 shows how these effects differ by school level. The order is sorted by the extent of the cream-skimming effects in matched TTA schools. If a comparison school is near a better cream-skimmed TTA school, the comparison school is more likely to suffer from greater negative impacts on new incoming left-behind students.

Those heterogeneous effects for new incoming students in TTA schools and matched comparison schools can have different cross-grade spillover effects for their older incumbent students. We first divide schools into two groups based on cream-skimming effects in TTA schools. Specifically, the first group (Group A) consists of TTA schools which had greater cream-skimming effects, and their matched comparison schools. The second group (Group B) consists TTA schools with less cream-skimming effects, and their matched comparison schools. The heterogeneous spillover effects by group level are estimated by adding interaction terms between each group dummy and main covariates in equations (2.1) and (2.2).

Table 2.7 shows the spillover effects in TTA schools. Since we divide schools into two groups based on the magnitudes of the heterogeneous cream-skimming effects, it is reasonable that Group A had greater increase in test score for newly selected students after the transition than that of Group B. However, even though Group A had more than twice greater cream-skimming effects than Group B, there is no significant cross-grade spillover effects in both groups.

Table 2.8 shows the spillover effects in matched comparison schools. As we mentioned in the previous section, there exist heterogeneous effects for new incoming left-behind students based on their groups. Only Group A suffered statistically significant left-behind effects in the average test scores and standard deviations. Matched comparison schools in Group A experienced 5.03% decrease in test scores and 4.74% increase in standard deviations. These changes in the distribution of newly admitted left-behind students had cross-grade spillover effects on incumbent students with 2.32% decrease in test scores and 3.12% increase in standard deviations. It is notable that even 16.92% increase in test scores of newly selected students in TTA schools did not have significant spillover effects across grades, whereas a decrease of 5.03% in test scores of newly admitted left-behind students in comparison schools had statistically significant spillover effects across grades.

4.4 Student Mobility

We then examined the changes in the pattern of student transfers since the introduction of autonomous schools. It is important to check the pattern of student transfers because the estimated effects in the previous subsections would be subject to bias if there is endogenous student mobility. For example, suppose that better incumbent students in non-autonomous schools who are mixed with lower-achieving incoming students, selectively transfer out. In this case, the estimated spillover effect in Equation (2.2) should be an underestimate of the true effect.

We believe that this kind of bias should not be severe in the context because student mobility is very limited due to the equalization policy. Under the policy, students cannot transfer to a different school in the same school district. It is quite costly to move to another district; furthermore, it is nearly impossible for high school students to transfer to another district, especially a better one, because of the minimum residential requirement. However, we check this concern empirically by estimating the following transfer equation:

$$\text{Transfer}_{gi't} = \beta \cdot LeftBehind_{gi't} + \delta \cdot PEER_{gi't} + \alpha_g + \alpha_{i'} + \alpha_t + \epsilon_{git} \quad (2.3)$$

where the dependent variable is a measure of student transfers, such as the number of transfer outs, the number of transfer ins, the transfer-out rate, and the transfer-in rate. We estimated the above equation for non-autonomous schools. Table 2.9 shows there were no significant transfers in matched comparison schools for newly admitted left-behind students and current students.

5 CONCLUSION

In this paper, we estimate spillover effects arising from the introduction of autonomous high schools. The findings indicate autonomous schools attracted high-achieving students. However, this influx of high-achieving students had no significant across-grade spillover effects on incumbent students in different grades who had entered the school before its autonomous transition. On the other hand, non-autonomous schools, especially those in close geographic proximity to autonomous ones, lost their high-achieving students to them due to cream-skimming. This higher proportion of lower-achieving students negatively affected the older incumbent students who had entered the school before the establishment of a nearby autonomous school.

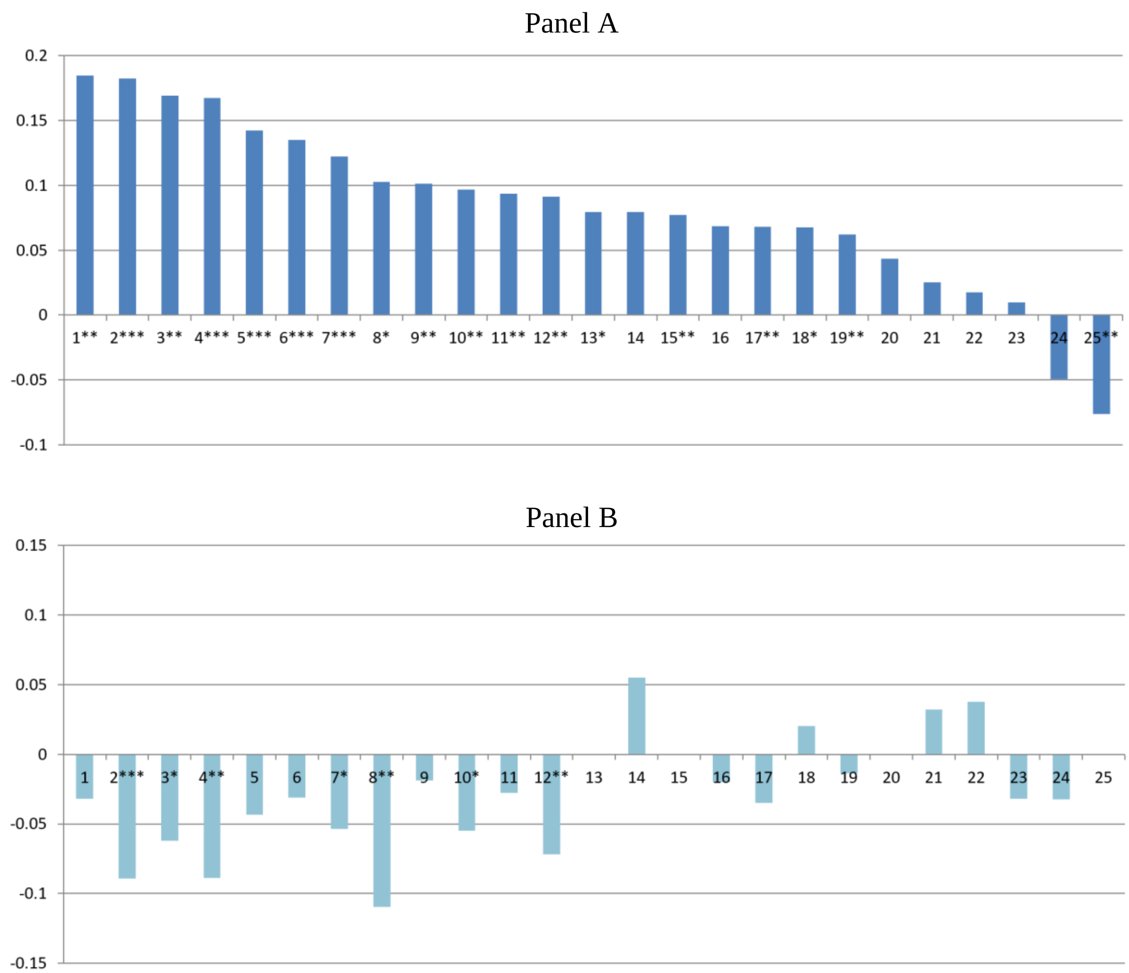
A possible interpretation is that any spillover effect from a positive change in student quality is insignificant or not large enough to offset, if any, a negative effect of school resource reallocation for accommodating new high-achieving and expensive students. In matched non-autonomous schools, there might not be any school resource reallocation in response to a negative change in student quality, or it is not large enough to offset the negative peer effect across grades. This is potentially important for educational policy making, as it suggests schools are not adequately flexible to respond to change in student quality.

Figure 2.1: Three Different Treated and Control Groups in a School

Entrance year	t-2	t-1	t	t+1	t+2	t+3
Before change (t-3)	**11**	**12**				
Before change (t-2)	**10**	**11**	**12**			
Before change (t-1)		**10**	**11**	**12**		
After change t			**10**	**11**	**12**	
After change (t+1)				**10**	**11**	**12**
After change (t+2)					**10**	**11**

Notes: Year t indicates the transition year of a TTA (or matched) school. Each column indicates an academic year and each low indicates an entrance year of a cohort. The number in a cell indicates grade. The control group is indicated as gray color, the group of new incoming students is indicated as red color, and the incumbent students of cross-grade spillover effect group is indicated as blue color.

Figure 2.2: Heterogeneous Cream Skimming Effects and Matched Left-Behind Effects



Note: The orders for TTA schools and the matched comparison schools are sorted by the magnitude of the cream-skimming effects in TTA schools. The 15th , 20th , and 25th TTA schools do not have matched schools. The left-behind estimate for the 13th comparison school is -0.0006. *** 1% significant; ** 5% significant; * 10% significant.

Table 2.1: Summary Statistics

	TTA schools			*Comparison schools*		
	Before	After		Before	After	
		New	Incumbent		New	Incumbent
Mean (log)	3.988	4.191	3.960	3.991	4.052	3.956
	(0.27)	(0.19)	(0.28)	(0.27)	(0.26)	(0.29)
Standard deviation (log)	2.941	2.687	2.927	2.956	2.870	2.957
	(0.38)	(0.42)	(0.38)	(0.36)	(0.36)	(0.34)
Credits per course	2.81	3.44	3.02	2.79	3.28	2.96
	(1.11)	(1.48)	(1.20)	(1.09)	(1.24)	(1.14)
Teaching hours per teacher	17.71	16.45		17.50	16.85	
	(1.09)	(2.31)		(0.95)	(1.48)	
School budget per pupil	51.63	66.09		33.77	38.06	
	(12.27)	(14.17)		(17.90)	(18.52)	
Temporary teachers	9.95	9.68		7.42	9.32	
	(5.24)	(4.45)		(6.23)	(4.95)	
Total number of teacher	75.59	72.86		75.81	74.68	
	(13.49)	(12.09)		(15.74)	(15.98)	
Year 2009	0.68	--	--	0.72	--	--
	(0.47)			(0.45)		
Year 2010	0.32	0.1	0.34	0.28	0.1	0.39
	(0.47)	(0.30)	(0.47)	(0.45)	(0.30)	(0.49)
Year 2011	--	0.33	0.49	--	0.33	0.47
		(0.47)	(0.50)		(0.47)	(0.50)
Year 2012	--	0.57	0.17	--	0.57	0.14
		(0.49)	(0.38)		(0.50)	(0.34)
Number of schools	25	25	25	22	22	22
School-year-course obs.	4476	3488	3535	3899	3483	3243

Notes: Standard deviations are presented in parentheses. Teaching hours per teacher is the average number of teaching hours per week. The unit of school budget per pupil is 1,000 KRW.

Table 2.2: Pre-transition Differences between TTA and Comparison Schools

	(1)	(2)
Credits	-0.021	-0.1687
	(0.5074)	(0.5447)
ln(Teaching hours per teacher)	0.0148	0.0056
	(0.0749)	(0.0785)
ln(school budget per student)	0.0169***	0.0174***
	(0.0040)	(0.0046)
Temporary teachers	-0.0004	0.0006
	(0.0149)	(0.0156)
Total number of teachers	0.0023	0.0026
	(0.0049)	(0.0049)
ln(mean score)		0.3153
		(1.0619)
ln(standard deviation)		-0.5934
		(0.9268)
Constant	-0.5762	0.4652
	(1.6351)	(5.8905)
Observations	47	47
R-squared	0.2795	0.292

Notes: Robust standard errors are presented in parentheses. The dependent variable is the indicator of whether the school is changed to autonomous school. Linear probability model. *** 1% significant; ** 5% significant; * 10% significant.

Table 2.3: Spillover Effects in TTA Schools: Means and Standard Deviations

	ln(Mean)		ln(Standard Deviation)	
	(1)	(2)	(3)	(4)
TTA	0.1692***	0.1759***	-0.2399***	-0.2330***
	(0.0135)	(0.0139)	(0.0260)	(0.0238)
PEER	-0.0099	-0.0034	-0.0159	-0.0086
	(0.0116)	(0.0126)	(0.0133)	(0.0162)
Credits		0.0072**		0.0042
		(0.0033)		(0.0049)
ln(Teaching hours per teacher)		-0.0047		-0.0754*
		(0.0183)		(0.0387)
ln(School budget per student)		-0.0411*		-0.0298
		(0.0211)		(0.0442)
Temporary teachers		0.0017*		0.0062***
		(0.0009)		(0.0020)
Total number of teachers		-0.0016**		-0.0052***
		(0.0007)		(0.0016)
Second semester	-0.1082***	-0.1081***	0.0527***	0.0527***
	(0.0038)	(0.0038)	(0.0039)	(0.0039)
11^{th} grade	-0.0121**	-0.0107*	0.0143	0.0143
	(0.0058)	(0.0058)	(0.0103)	(0.0104)
12^{th} grade	-0.1432***	-0.1411***	0.0252*	0.0258*
	(0.0089)	(0.0090)	(0.0128)	(0.0132)
2010	-0.0131**	-0.0153**	-0.0016	-0.0045
	(0.0062)	(0.0063)	(0.0078)	(0.0095)
2011	0.0698***	0.0705***	-0.0837***	-0.0824***
	(0.0116)	(0.0117)	(0.0134)	(0.0129)
2012	0.0255**	0.0240*	0.0138	-0.0055
	(0.0126)	(0.0132)	(0.0156)	(0.0169)
Constant	4.1074***	4.3611***	2.9030***	3.5539***
	(0.0072)	(0.1321)	(0.0119)	(0.3352)
Course-by-school Effects	Yes	Yes	Yes	Yes
Observations	11,499	11,499	11,499	11,499
R-squared	0.6781	0.6793	0.7270	0.7291

Notes: Robust standard errors, clustered by school-course, are presented in parentheses. The dependent variable is log mean score for Column (1) and (2) and log standard deviation for (3) and (4). TTA refers students admitted after transition. Peer refers incumbent students. *** 1% significant; ** 5% significant; * 10% significant.

Table 2.4: Spillover Effects in Matched Schools: Means and Standard Deviations

	ln(Mean)		ln(Standard Deviation)	
	(1)	(2)	(3)	(4)
Left-behind	-0.0318***	-0.0376***	0.0373**	0.0374**
	(0.0109)	(0.0107)	(0.0157)	(0.0150)
PEER	-0.0133	-0.0165*	0.0249**	0.0272**
	(0.0094)	(0.0097)	(0.0116)	(0.0116)
Credits		0.0108***		0.0154***
		(0.0036)		(0.0052)
ln(Teaching hours per teacher)		-0.0226		0.0238
		(0.0303)		(0.0324)
ln(School budget per student)		0.0094		-0.0251
		(0.0192)		(0.0204)
Temporary teachers		0.0005		-0.0014
		(0.0012)		(0.0015)
Total number of teachers		0.0009		0.0000
		(0.0011)		(0.0014)
Second semester	-0.1206***	-0.1205***	0.0273***	0.0274***
	(0.0048)	(0.0049)	(0.0043)	(0.0044)
11^{th} grade	-0.0088	-0.0068	0.0154*	0.0186**
	(0.0062)	(0.0062)	(0.0093)	(0.0094)
12^{th} grade	-0.1605***	-0.1577***	0.0188*	0.0235**
	(0.0105)	(0.0106)	(0.0112)	(0.0114)
2010	-0.002	-0.0022	-0.0048	-0.0072
	(0.0074)	(0.0073)	(0.0074)	(0.0084)
2011	0.0827***	0.0814***	-0.1125***	-0.1117***
	(0.0112)	(0.0125)	(0.0149)	(0.0146)
2012	0.0103	0.004	-0.0628***	-0.0612***
	(0.0130)	(0.0147)	(0.0181)	(0.0166)
Constant	4.1228***	4.0517***	2.9248***	2.9027***
	(0.0081)	(0.1572)	(0.0093)	(0.1783)
Course-by-school Effects	Yes	Yes	Yes	Yes
Observations	10,625	10,625	10,625	10,625
R-squared	0.6376	0.6391	0.685	0.6868

Notes: Robust standard errors, clustered by school-course, are presented in parentheses. The dependent variable is log mean score for Column (1) and (2) and log standard deviation for (3) and (4). *** 1% significant; ** 5% significant; * 10% significant.

Table 2.5: Average Cream-Skimming Effects in TTA Schools: Means and Standard Deviations

	ln(Mean)		ln(Standard Deviation)	
	(1)	(2)	(3)	(4)
TTA	0.0824***	0.0884***	-0.1976***	-0.2212***
	(0.0142)	(0.0149)	(0.0267)	(0.0278)
Credits		0.0084		0.0024
		(0.0068)		(0.0137)
ln(Teaching hours per teacher)		0.0344		-0.0168
		(0.0291)		(0.0549)
ln(School budget per student)		-0.0405		0.0493
		(0.0307)		(0.0549)
Temporary teachers		0.0007		0.0022
		(0.0016)		(0.0026)
Total number of teachers		-0.0003		-0.0078***
		(0.0010)		(0.0022)
2010	0.0174*	0.0157	-0.0525***	-0.0394**
	(0.0103)	(0.0108)	(0.0160)	(0.0169)
2011	0.0987***	0.0924***	-0.1465***	-0.1260***
	(0.0148)	(0.0167)	(0.0258)	(0.0301)
2012	0.0613***	0.0633***	-0.0376	-0.0495
	(0.0150)	(0.0171)	(0.0254)	(0.0323)
Constant	4.1150***	4.1688***	2.8325***	3.2393***
	(0.0066)	(0.1828)	(0.0113)	(0.3273)
Course-by-school Effects	Yes	Yes	Yes	Yes
Observations	1,170	1,170	1,170	1,170
R-squared	0.7887	0.7906	0.8581	0.8631

Notes: Robust standard errors, clustered by school-course, are presented in parentheses. The dependent variable is log mean score for Column (1) and (2) and log standard deviation for (3) and (4). *** 1% significant; ** 5% significant; * 10% significant.

Table 2.6: Average Left-Behind Effects in Matched Schools: Means and Standard Deviations

	ln(Mean)		ln(Standard Deviation)	
	(1)	(2)	(3)	(4)
Left-behind	-0.0309*	-0.0310*	0.0684*	0.0574**
	(0.0173)	(0.0178)	(0.0363)	(0.0265)
Credits		-0.0123**		0.0309
		(0.0052)		(0.0235)
ln(Teaching hours per teacher)		-0.0512		0.0308
		(0.0407)		(0.0732)
ln(School budget per student)		0.0212		-0.0535
		(0.0295)		(0.0382)
Temporary teachers		-0.0009		-0.0035
		(0.0018)		(0.0044)
Total number of teachers		-0.0006		0.0042
		(0.0020)		(0.0031)
2010	-0.0117	-0.0102	-0.011	-0.0104
	(0.0126)	(0.0141)	(0.0178)	(0.0197)
2011	0.1131***	0.1251***	-0.1776***	-0.1783***
	(0.0193)	(0.0239)	(0.0474)	(0.0441)
2012	0.0394*	0.0486**	-0.0762	-0.0694
	(0.0204)	(0.0245)	(0.0507)	(0.0440)
Constant	4.1416***	4.3025***	2.7892***	2.4910***
	(0.0069)	(0.2457)	(0.0091)	(0.3537)
Course-by-school Effects	Yes	Yes	Yes	Yes
Observations	1,088	1,088	1,088	1,088
R-squared	0.7765	0.7788	0.8504	0.8551

Notes: Robust standard errors, clustered by school-course, are presented in parentheses. The dependent variable is log mean score for Column (1) and (2) and log standard deviation for (3) and (4). *** 1% significant; ** 5% significant; * 10% significant.

Table 2.7: Spillover Effects in TTA Schools: Means and Standard Deviations

	ln(Mean)		ln(Standard Deviation)	
	(1)	(2)	(3)	(4)
GroupA*TTA	0.1692***	0.1759***	-0.2399***	-0.2330***
	(0.0135)	(0.0139)	(0.0260)	(0.0238)
GroupA*PEER	-0.0099	-0.0034	-0.0159	-0.0086
	(0.0116)	(0.0126)	(0.0133)	(0.0162)
GroupB*TTA	0.0801***	0.0822***	-0.1944***	-0.1993***
	(0.0127)	(0.0128)	(0.0187)	(0.0204)
GroupB*PEER	-0.0168	-0.0145	0.0048	-0.0007
	(0.0107)	(0.0112)	(0.0138)	(0.0148)
Credits		0.0072**		0.0042
		(0.0033)		(0.0049)
ln(Teaching hours per teacher)		-0.0047		-0.0754*
		(0.0183)		(0.0387)
ln(School budget per student)		-0.0411*		-0.0298
		(0.0211)		(0.0442)
Temporary teachers		0.0017*		0.0062***
		(0.0009)		(0.0020)
Total number of teachers		-0.0016**		-0.0052***
		(0.0007)		(0.0016)
Second semester	-0.1082***	-0.1081***	0.0527***	0.0527***
	(0.0038)	(0.0038)	(0.0039)	(0.0039)
11^{th} grade	-0.0121**	-0.0107*	0.0143	0.0143
	(0.0058)	(0.0058)	(0.0103)	(0.0104)
12^{th} grade	-0.1432***	-0.1411***	0.0252*	0.0258*
	(0.0089)	(0.0090)	(0.0128)	(0.0132)
2010	-0.0131**	-0.0153**	-0.0016	-0.0045
	(0.0062)	(0.0063)	(0.0078)	(0.0095)
2011	0.0698***	0.0705***	-0.0837***	-0.0824***
	(0.0116)	(0.0117)	(0.0134)	(0.0129)
2012	0.0255**	0.0240*	0.0138	-0.0055
	(0.0126)	(0.0132)	(0.0156)	(0.0169)
Constant	4.1074***	4.3611***	2.9030***	3.5539***
	(0.0072)	(0.1321)	(0.0119)	(0.3352)
Course-by-school Effects	Yes	Yes	Yes	Yes
Observations	11,499	11,499	11,499	11,499
R-squared	0.6781	0.6793	0.7270	0.7291

Notes: Robust standard errors, clustered by school-course, are presented in parentheses. The dependent variable is log mean score for Column (1) and (2) and log standard deviation for (3) and (4). *** 1% significant; ** 5% significant; * 10% significant.

Table 2.8: Spillover Effects in matched Schools: Means and Standard Deviations

	ln(Mean)		ln(Standard Deviation)	
	(1)	(2)	(3)	(4)
GroupA*Left-behind	-0.0503***	-0.0565***	0.0474***	0.0502***
	(0.0115)	(0.0118)	(0.0159)	(0.0166)
GroupA*PEER	-0.0232**	-0.0268**	0.0312**	0.0343**
	(0.0105)	(0.0108)	(0.0129)	(0.0134)
GroupB*Left-behind	-0.0061	-0.009	0.0232	0.0179
	(0.0133)	(0.0134)	(0.0197)	(0.0182)
GroupB*PEER	-0.0037	-0.0051	0.0183	0.0191
	(0.0121)	(0.0121)	(0.0151)	(0.0148)
Credits		0.0106***		0.0155***
		(0.0036)		(0.0053)
ln(Teaching hours per teacher)		-0.0002		0.0087
		(0.0300)		(0.0331)
ln(School budget per student)		0.0201		-0.0324
		(0.0194)		(0.0217)
Temporary teachers		0.0007		-0.0016
		(0.0012)		(0.0015)
Total number of teachers		-0.0005		0.0009
		(0.0011)		(0.0014)
Second semester	-0.1205***	-0.1204***	0.0272***	0.0273***
	(0.0049)	(0.0049)	(0.0043)	(0.0043)
11^{th} grade	-0.0085	-0.0064	0.0153	0.0183*
	(0.0061)	(0.0062)	(0.0093)	(0.0094)
12^{th} grade	-0.1598***	-0.1568***	0.0185*	0.0228**
	(0.0105)	(0.0106)	(0.0111)	(0.0113)
2010	0.0001	0.0031	-0.0061	-0.0108
	(0.0072)	(0.0071)	(0.0075)	(0.0089)
2011	0.0846***	0.0806***	-0.1135***	-0.1112***
	(0.0113)	(0.0126)	(0.0149)	(0.0146)
2012	0.0117	0.0039	-0.0634***	-0.0611***
	(0.0132)	(0.0148)	(0.0180)	(0.0166)
Constant	4.1224***	4.0577***	2.9251***	2.8978***
	(0.0082)	(0.1480)	(0.0093)	(0.1777)
Course-by-school Effects	Yes	Yes	Yes	Yes
Observations	10,625	10,625	10,625	10,625
R-squared	0.6386	0.6400	0.6851	0.6870

Notes: Robust standard errors, clustered by school-course, are presented in parentheses. The dependent variable is log mean score for Column (1) and (2) and log standard deviation for (3) and (4). *** 1% significant; ** 5% significant; * 10% significant.

Table 2.9: Student Transfers in Matched Schools

Matched Schools	(1) Out	(2) Out+Quit	(3) In	(4) Out	(5) Out+Quit	(6) In
Left-behind	0.0861	1.8158	2.5201	-0.001	0.0034	0.0070*
	(1.4848)	(2.3485)	(1.9858)	(0.0033)	(0.0048)	(0.0041)
PEER	0.483	0.4187	0.8318	0.0001	0.0006	0.0027
	(1.2672)	(2.0042)	(1.6947)	(0.0028)	(0.0041)	(0.0035)
11th grade	-10.6552***	-16.3988***	-11.9356***	-0.0242***	-0.0364***	-0.0260***
	(0.6323)	(1.0000)	(0.8456)	(0.0014)	(0.0020)	(0.0018)
12th grade	-16.9988***	-29.0732***	-18.6669***	-0.0388***	-0.0649***	-0.0410***
	(0.7151)	(1.1311)	(0.9564)	(0.0016)	(0.0023)	(0.0020)
Year 2009	-1.2727	1.1818	1.7727*	-0.0018	0.0039	0.0044**
	(0.7832)	(1.2387)	(1.0474)	(0.0018)	(0.0025)	(0.0022)
Year 2010	-2.5709**	0.1895	-1.8847	-0.0041*	0.0015	-0.0041
	(1.0837)	(1.7141)	(1.4494)	(0.0024)	(0.0035)	(0.0030)
Year 2011	-1.8937	0.1131	-0.9999	-0.0024	0.0015	-0.0025
	(1.5213)	(2.4062)	(2.0346)	(0.0034)	(0.0049)	(0.0042)
Year 2012	-1.5571	-1.6268	-2.5697	-0.0011	-0.0017	-0.0054
	(1.6215)	(2.5647)	(2.1686)	(0.0036)	(0.0052)	(0.0045)
Constant	18.9373***	31.0670***	19.8955***	0.0422***	0.0685***	0.0432***
	(0.6812)	(1.0775)	(0.9111)	(0.0015)	(0.0022)	(0.0019)
School Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	333	333	333	333	333	333
R-squared	0.7444	0.7855	0.7032	0.7445	0.8037	0.7110

Notes: 22 matched schools. Out= transfer-outs, In = transfer-ins, Quit = quits. The dependent variable is the number of students in Columns (1)-(3) and the rate of transfer in Columns (4)-(6). Robust standard errors, clustered by school, are presented in parentheses. *** 1% significant; ** 5% significant; * 10% significant.

Chapter 3: Learning Knowledge or Test Taking Skill? The Impact of the Number of Tests on Students Academic Performance

1 INTRODUCTION

In the past few years, there has been substantial concern among parents and policy-makers that we are doing too much testing of students (Zernike 2015). These excessive tests were further exacerbated since pass of the No Child Left Behind Act from Congress in 2001, which stressed on school accountability that aims to overcome the principal-agent problem by incentivizing educators to focus more on student achievement. In other words, as reported measured school performance can lead to positive or negative consequences based on their performance, educators have incentives for better measured performance such as concentrating on the materials that are being measured. However, if educators are too incentivized, school accountability might even hinder student learning by reducing the breadth of material teachers cover; this is referred to as “teaching to the test”. Furthermore, larger number of testing could crowd out time otherwise used for learning or opportunities at school.

A few years ago, former President Obama responded to the concerns of over-testing students while he was in office, and mentioned “in moderation, smart, strategic tests” are still needed (Doering 2015). In his statement, the concern over teaching to the test is addressed as “smart and strategic” tests, and the concern over excessive number of tests is addressed as “moderate” tests. We already know the importance of considering the specific nature of assessments when we decide how broadly to base an accountability system to account for teaching to the test even though it is not simple in practice (Figlio and Loeb 2010, 390). However, it is unclear whether modestly relaxing the number of tests would be beneficial especially when teachers’ behavior or teaching is controlled for. We can expect a greater number of tests could improve student performance by incentivizing their time spent studying and by improving test-taking skills. Exams could also hinder student learning by encouraging students to focus too heavily on the material and techniques covered on the test. Figlio and Loeb (2010) review school accountability, that inevitably includes more tests, literature and finds some positive effects of accountability on mathematics test scores. However, the impact of school accountability inevitably contains much of educators' behavior since school accountability is intended to motivate educators' teaching and behaviors for better student performance.

In this study, I examine the impact of eliminating two exams in one year on later test scores exploiting a phased-in policy in Korea in which students are exempt from second-semester exams in their first year in middle school without much changes in teachers' behavior with a difference-in-differences strategy.

This natural experiment in South Korea is unique because there are not enough reasons or incentives for educators to change their behaviors. Even when there were midterm and finals, those tests were not used for accountability which suggests there was no incentive for educators to improve students' test scores beyond their normal teaching duties, and this can be applied to in the same way when students are exempt from the tests. Moreover, total academic class time has remained unchanged as schools utilize their initial exam days for extra-curricular activities. Korea also provides a good setting in which to examine the causal impact of the number of tests on later academic performance because there exists no selection bias due to an equalization policy. Students do not have choice as to which school to attend; instead, they are assigned to schools within school districts by the regional office of education. Moreover, strictly regulated curriculum and tuition under the equalization policy resulted in similar schools across regions. All of these factors make it more likely that I am identifying the casual impact of less tests on later academic performance.

I find that taking two tests less decreased middle school math test scores in next school year by 0.053 standard deviations. However, there was no significant impact on middle school reading test scores. This result is interesting because taking less tests seems to have an impact on test scores in the following school year, especially in math but not in reading.

The rest of the paper proceeds as follows. Section 2 presents the institutional details about the natural experiment. Section 3 describes the data sets used in this study and then presents the estimation strategies. Section 4 presents the results including a robustness check, and Section 5 concludes.

2 BACKGROUND

In South Korea, primary education is completed in elementary school, from 1st to 6th grade. Secondary education starts in middle school, from 7th to 9th grade, and ends in high school, 10th to 12th grade. Although compulsory education ends with middle school, nearly all students attend an academic or a vocational high school. An academic year coincides with a calendar year in South Korea; in other words, the first semester begins in March (Spring semester) and then the second begins in September (Fall semester).

Middle school students take midterm and final exams every semester, and their test scores are important for students.[31] Students and their parents can choose either a vocational high school or academic high school based on the students' interest and test scores. Moreover, their test scores are important for admittance into special-purpose high schools and autonomous high schools, which are not subject to the equalization policy.

This study examines the causal impact of testing students less on their academic performance. I use a difference-in-differences design to exploit a natural experiment in

31 On the other hand, there are no incentives or penalties for teachers based on those test scores.

South Korea in which some schools introduced the "free semester" for 7th graders earlier than the others. Under this policy, 7th grade students were not required to take mid-term or final exams in their second semester, and resumed to take them in their subsequent school years.[32]

The "free semester" program was an election pledge of the former president, Geun-hye Park, during her successful 2012 campaign. The government introduced the policy suddenly in 2013 and schools were required to adopt it by 2016. Initially, 42 middle schools adopted the policy for the 2013 academic year. That number then increased to 811 middle schools (26%) in 2014 and 2,230 middles schools (70%) in 2015. I focus on middle schools in Seoul because other regions in the Capital area adopted delayed school start time during the time period.[33] In Seoul, 1.3% of middle schools adopted the policy in 2013, and an additional 37.6% adopted it in 2014. Thus, the remaining 61.1% of middle schools adopted the policy after 2014.

The natural experiment in South Korea is unique because there were no incentives for educators to change their behaviors during the phased-in period. There were no standardized tests for middle school students until 9th grade; therefore, students only had midterms and finals in their 7th and 8th grades, and those test results were not used for accountability. In other words, teachers had no incentive to improve their students' test performance, and to go beyond their normal teaching duties. Therefore, even when students are exempt from tests in their second semester in 7th grade, teachers had no incentive to change their behaviors, especially in teaching. Moreover, the policy did not increase total

---

32 Initially, schools could choose which semester to be exempt from the tests among any one semester in 7th and 8th grades, but most schools chose the second semester in 7th grade since 2014.

33 Even though Seoul cover 0.6% of South Korea, about 20% of total population lived in Seoul in 2015.

academic class time since students enjoyed additional extracurricular activities during their initial exam days.

The educational system in Korea also provides a good setting in which to examine the causal impact of the number of tests on later academic performance due to an equalization policy. First of all, under an equalization policy, students do not have choice as to which school to attend because they were assigned to schools within school districts by the regional office of education. This means that there is no selection bias at least on a student side. Moreover, under an equalization policy, curriculum and tuition are strictly regulated in a national level to insure very little difference across regions and schools

3 DATA AND ESTIMATION STRATEGY

3.1 Data

The main data set in this study is derived from an online school information reporting system. The Korean Education & Research Information Service (KERIS) collects the school level information, including test scores. I collected data from 2013 to 2015 for all middle schools in Seoul.[34]

Academic performance reported in the data set is the mean and standard deviation of students' test scores in each course by school, grade, year, and semester level. The data also reveals the number of students in each school and grade so that I can normalize each school grade score as student level normalized data by weighting the number of students in each school and grade. Although each school produces their own exams, nationally standardized curriculum and textbooks render the scores comparable across schools.[35]

---

34 I focus on middle school students' test scores in Seoul because other regions in the Capital area adopted delayed school start time during this time period. (See Chapter 1)

35 All schools in South Korea can use a textbook that are approved by the Ministry of Education

Table 3.1 presents the number of cumulative tests that students had taken in each semester in treatment (Panel A) and control (Panel B) schools. Since there is no testing during primary education in Seoul, middle school students took their first and second tests during Spring semester, and the third and fourth tests in Fall semester in 7th grade. Subsequently, they took their fifth and sixth tests during Spring semester, and seventh and eighth tests during Fall semester in the following school year when they advanced to the 8th grade. However, as midterm and finals were exempted in the second semester in 2014, students in the 7th grade had their third and fourth tests in their Spring semester, and fifth and sixth tests in Fall semester in the following school year of 2015 in treatment schools. Therefore, the number of tests 8th grade students had taken are less by two in treatment schools compared to control schools in 2015, and I focus on 8th grade students' test scores.

Table 3.2 shows 8th grade students' average test scores for the pre- and post-periods in Seoul by their timing of adopting "free semester". There are 378 middle schools in the data set, wherein 144 schools (38%) implemented the "free semester" program in 2014, and the other 234 schools (62%) adopted it after 2014.[36] Math test scores show larger difference between treatment and control groups in pre-period which is statistically significant at the 10% level; however, in the post-period, the difference becomes even the opposite. This results in a 1.21-point decrease in math which is statistically significant at the 1% level. On the other hand, reading test scores show almost no differences between treatment and control groups in both pre- and post-treatment periods.

---

[36] There are 383 middle school in Seoul, but 5 schools were dropped from the sample because they started to implement the policy in 2013.

3.2 Estimation Strategy

To measure the causal effect testing students less has on academic performance, I estimate the following equation weighted by the number of students in each school to generates student-level estimates:

$$Y_{sdcyt} = \beta_0 + \beta_1 treatment_s * post_y + \beta_2 treatment_s + \beta_3 post_y \qquad (3.1)$$
$$+ \mu X_s + \gamma_d + \delta_c + \omega_y + \varphi_t + \varepsilon_{sdcyt}$$

Y_{sdcyt} is a dependent variable for school s in school district d, and subject c, at year y and semester t. The variable $treatment_s$ is an indicator equal to 1 if school s adopted the "free semester" earlier in 2014 and 0 otherwise. The variable $post_y$ is an indicator equal to 1 if the dependent variable was observed in 2015 and 0 otherwise. Therefore, the key coefficient of interest is β_1, which measures the difference-in-differences estimate. A vector of school-specific control variables X_s includes whether the school is a private school, single-sex male school, and single-sex female school or not.[37] γ_d is a school a school district fixed effect, δ_c is a course fixed effect, ω_y is a year fixed effect, and φ_t is a semester fixed effect, indicating whether it is fall semester or not.

The key assumption underlying a difference-in-differences strategy is that the outcome of interest would follow the same trend in the absence of the treatment. Although I cannot test this assumption directly, I can examine whether the pre-treatment trends are the same. Figures 3.1 and 3.2 present the trends in the 8th grade students' test scores by semester in treatment and control schools for math and reading, respectively. Parallel time trends in the pre-period can be found in both subjects. Math test scores are greater in treatment schools in pre-period, but the order is revered after the treatment. Reading test scores are almost the same in both groups of schools.

[37] When I include a school fixed effect in later estimation, these time invariant school characteristic variables are omitted from the regression.

Balance test examines the differences between earlier and later "free semester" adopted schools. Table 3.3 presents the results with a linear probability model. Column (1) shows there was no significant difference between the treatment and control schools in terms of test scores, but Column (2) presents the existing differences in some of the school characteristics. Single-sex and public schools were more likely to adopt the policy earlier. Thus, these factors must be controlled for in later estimations. Nevertheless, further examinations are needed to check the possibilities of endogeneity.

4 RESULTS

4.1 Main Results

Table 3.4 shows the impact of the policy on students' test scores. Column (1) presents the baseline estimate of the difference-in-differences, while only controlling for course, year, and semester effects.[38] I find that the policy has a negative impact on later test scores, though it is statistically significant only at the 10 % level. This estimate is robust to the inclusion of a variety of control variables. Column (2) adds school characteristics such as private school and single-sex school effects, as these were significant at the balance test. Column (3) includes school district effects, and Column (4) replaces all the time invariant school related variables with school fixed effects. Taking two tests less decreased test scores by 0.028 standard deviations on average. In most cases, estimates are significant at the 10 % level. When school fixed effects are included, the point estimate is still consistent, which confirms that school specific characteristics do not cause bias in the results; however, including school fixed effects decreases the magnitude by a little, increase standard errors as well, and now the estimate is marginally insignificant.

---

[38] Even when year and semester fixed effects are excluded from the baseline estimation, the results are almost identical.

The impact of the policy on math test score is presented in Table 3.5, which is organized similarly to Table 3.4. According to the baseline estimate, students' math test scores decreased by 0.053 standard deviations on average after taking two tests less during the last school year. This statistically significant negative impact is robust to the inclusion of various control variables. Most of the estimates are statistically significant at the 1% level, and is still significant at the 5% level even when school fixed effects are included.

Table 3.6 presents the impact of the policy on reading test scores, organized in similar fashion as in Table 3.4. All the estimates show that the policy does not have a significant impact on reading test scores. This different effect on each subject is interesting because it suggests the number of tests does not impact test taking skills, unless math and reading required as different test taking skillset.

4.2 Robustness

Difference-in-differences requires parallel time trends and a good method in examining it is the inclusion of leads in the difference-in-differences model as Autor (2003) did. I included three leads and two lags in Equation (3.1) as below:

$$Y_{sdct} = \alpha_0 + \sum_{\tau=-3}^{-1} \beta_\tau D_{s\tau} + \sum_{\tau=1}^{2} \beta_\tau D_{s\tau} + \alpha_1 treatment_s + \varphi_t + \mu X_s + \gamma_d + \delta_c + \varepsilon_{sdct} \quad (3.2)$$

where τ is a relative timing with normalized to 0 at the time of the treatment, so a negative value indicates a lead. I also added lags to examine the post-treatment effects and their persistence. Parallel time trends require that leads have estimates which are close to zero and statistically insignificant.

Figures 3.3 and 3.4 present math and reading test score results by semester. Coefficient estimates of the leads and the lags are presented with 95% confidence intervals. In both figures, the confidence intervals of leads contain zero, which means leads are statistically insignificant at the 5% level, whereas the estimates of the lags have negative values and are statistically significant at the 5% level, only for math in Figure 3.3, and this remains relatively constant.

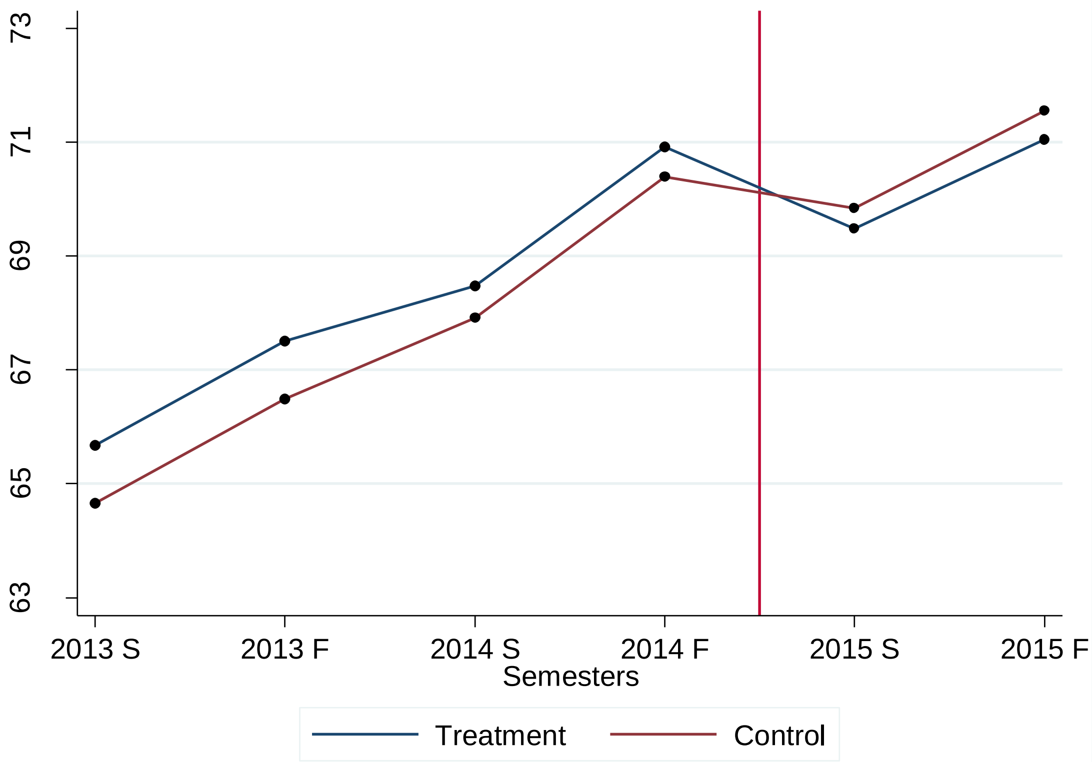
5 CONCLUSION

In this paper, I examine how the number of tests affects students' later test scores by exploiting a natural experiment in which students are exempt from midterm and final exams during a semester in 2014 only in some school in South Korea. I find that taking two tests less during the last school year of middle school decreased math test scores by 0.053 standard deviations, but does not have an impact on reading test scores. This negative impact on math test scores remains relatively constant for at least one academic year.

Since the "free semester" policy was intended to ease the students' excessive stress from tests, students' test scores may not be a good outcome to measure the impact of the policy. Although this study cannot answer the whole aspects of the policy, it is still worth studying since examined test scores are those of the following school years and this study can provide a better understanding on the number of tests.

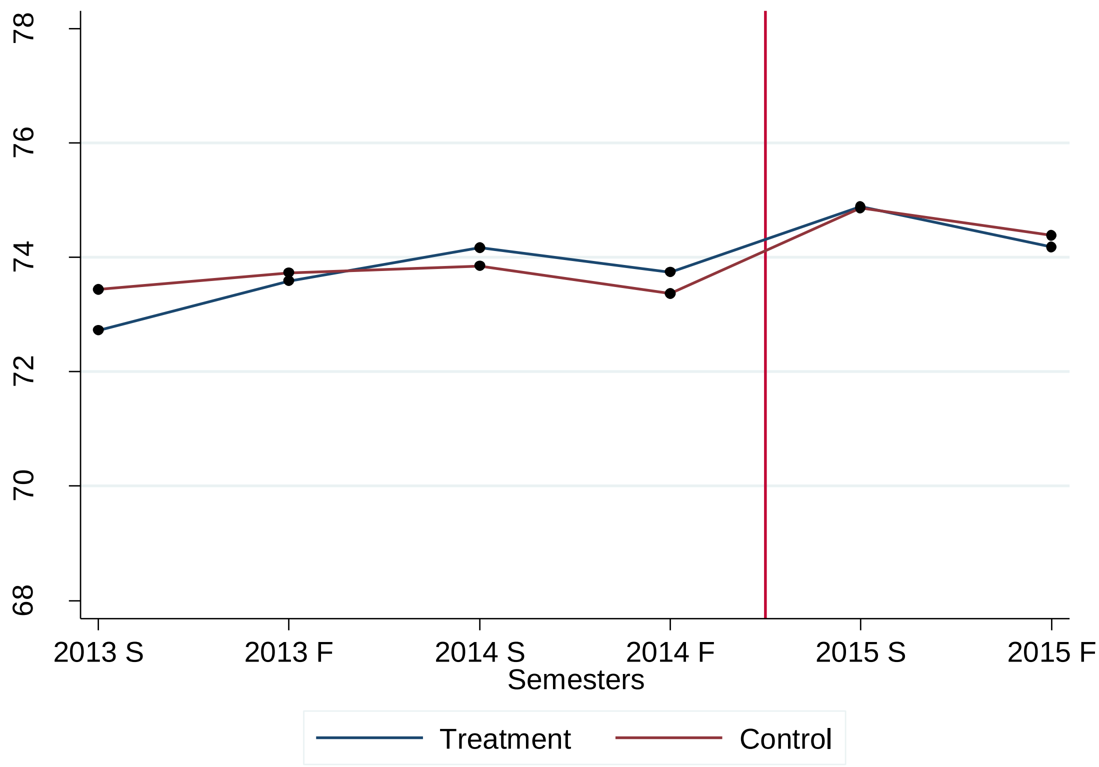
Overall, I find that the positive impacts of having more tests outperform the negative impacts. This means more tests can incentivize students to spend additional time studying and in learning more knowledge or test taking skills. However, it is notable that less tests do not have a harmful effect on reading, but on math. This suggests the number of tests has an impact on students' academic performance through learning more or less knowledge than test taking skills. When we compare reading and math, math requires more prior knowledge and is especially learned in class; however it is less likely that required test taking skills are different based on subject, as most portion of the tests in the data is multiple choice questions.

Figure 3.1: Trends in Math Test Scores of 8th Grade in Middle School



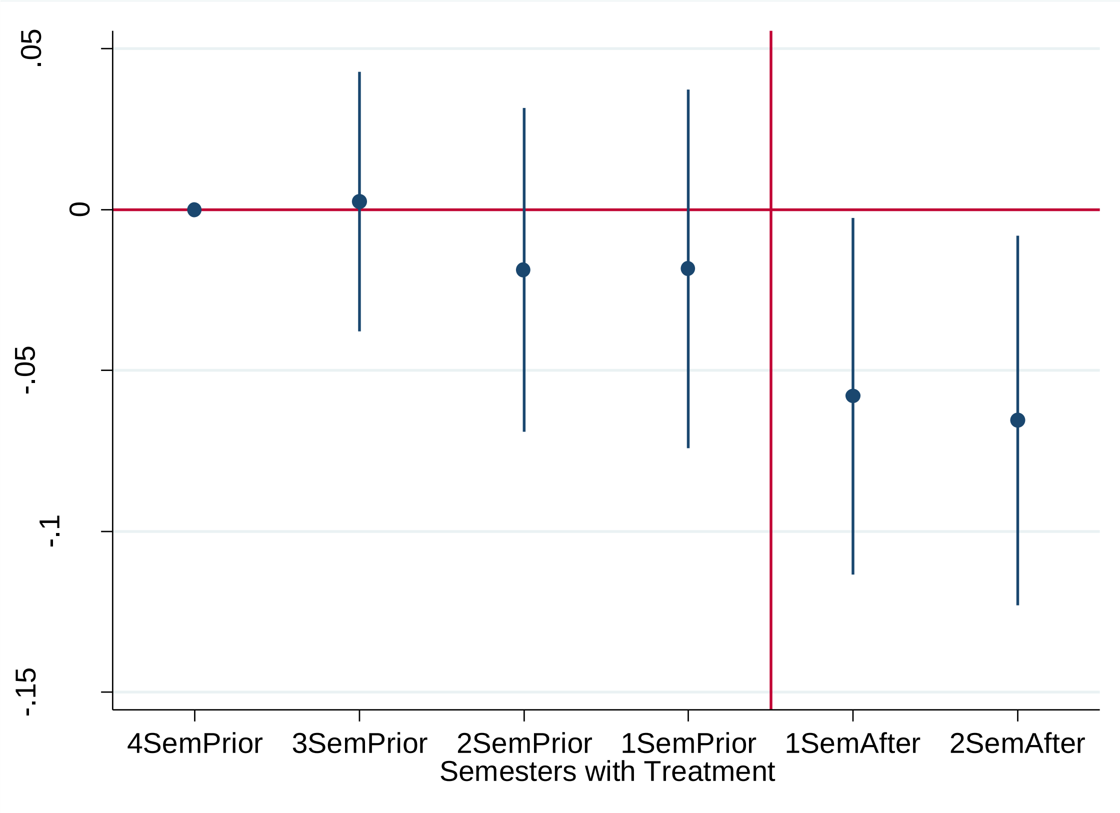
Notes: Average math test scores for 8th grade middle school students in all Seoul schools are presented from 2013 spring semester through 2015 fall semester. S refers spring semester and F refers fall semester.

Figure 3.2: Trends in Reading Test Scores of 8th Grade in Middle School



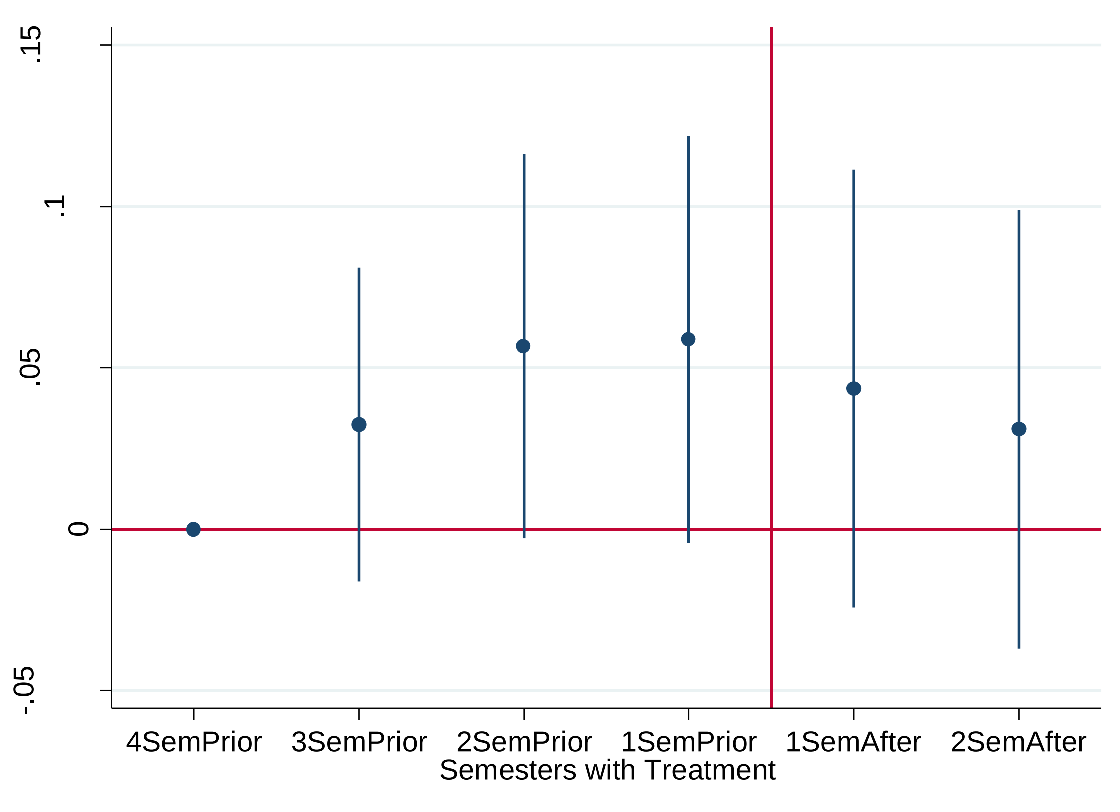
Notes: Average reading test scores for 8th grade middle school students in all Seoul schools are presented from 2013 spring semester through 2015 fall semester. S refers spring semester and F refers fall semester.

Figure 3.3: Estimated Impact of Free Semester on Test Scores by Semesters on Math



Notes: Estimated impact of free semester on math test scores is presented with 95% confidence intervals. The dependent variable is standardized score. Estimates are from a model that allows for effects before and after the experience of free semester. The base line effect is that of 4 semesters prior.

Figure 3.4: Estimated Impact of Free Semester on Test Scores by Semesters on Reading



Notes: Estimated impact of free semester on reading test scores is presented with 95% confidence intervals. The dependent variable is standardized score. Estimates are from a model that allows for effects before and after the experience of free semester. The base line effect is that of 4 semesters prior.

Table 3.1: The Number of Tests that Students Took in Each Semester by School

Panel A: The Number of Tests Students Took in Treatment School

Semester	Year 2013			Year 2014			Year 2015	
	Spring	Fall		Spring	Fall		Spring	Fall
7th Grader	1,2	3,4	↘	1,2	-	↘	1,2	-
8th Grader	5,6	7,8	↘	5,6	7,8	↘	3,4	5,6
9th Grader	9,10	11,12	↘	9,10	11,12	↘	9,10	11,12

Panel B: The Number of Tests Students Took in Control School

Semester	Year 2013			Year 2014			Year 2015	
	Spring	Fall		Spring	Fall		Spring	Fall
7th Grader	1,2	3,4	↘	1,2	3,4	↘	1,2	- (or 3,4)
8th Grader	5,6	7,8	↘	5,6	7,8	↘	5,6	7,8
9th Grader	9,10	11,12	↘	9,10	11,12	↘	9,10	11,12

Notes: The numbers indicate the cumulative number of tests students took in each semester.

Table 3.2: Average Test Scores with DID without Controls

Time	Pre			Post			DID	
Schools	Treatment (144 Schools)	Control (234 Schools)	Difference (A)	Treatment (144 Schools)	Control (234 Schools)	Difference (B)	DID (B-A)	t-stat
Reading	73.56	73.60	-0.04	74.54	74.63	-0.08	-0.04	0.11
Math	68.15	67.37	0.78	70.27	70.70	-0.43	-1.21	2.85
Observations	574	932		288	468			

Notes: Average test scores for reading and math are presented for pre- and post-period in treatment and control schools. Difference-in-differences of means are presented. Standard errors are clustered by school.

Table 3.3: Differences between earlier and later free-semester policy implementing schools

	(1)	(2)
Mean Score	0.002	-0.001
	(0.003)	(0.003)
Standard Deviation	0.006	-0.003
	(0.005)	(0.005)
Number of Students		-0.000
		(0.000)
Private School		-0.454***
		(0.032)
Single Sex School: Male		0.082**
		(0.040)
Single Sex School: Female		0.157***
		(0.042)
School District Dummies		yes
Constant	0.119	0.606**
	(0.304)	(0.288)
Observations	1,500	1,500
R-squared	0.001	0.153

Notes: Robust standard errors are presented in parentheses. The dependent variable is the indicator of whether the school implemented free-semester in 2014. Linear probability model. *** 1% significant; ** 5% significant; * 10% significant.

Table 3.4: The Effects of Free Semester on Middle School Standardized Test Scores

	(1)	(2)	(3)	(4)
VARIABLES	Standardized Average Test Score for Math and Reading			
Treated*Post	-0.028*	-0.027*	-0.028*	-0.025
	(0.014)	(0.014)	(0.014)	(0.015)
Treated	0.016	0.027*	0.024	0.024
	(0.014)	(0.015)	(0.015)	(0.015)
Private School		0.048*	0.057**	
		(0.028)	(0.028)	
Single Sex School: Male		-0.066**	-0.078**	
		(0.029)	(0.032)	
Single Sex School: Female		0.085***	0.070**	
		(0.030)	(0.032)	
School District Effects			✓	
School Effects				✓
Observations	4,524	4,524	4,524	4,524
R-squared	0.001	0.033	0.073	0.580

Notes: Standard errors, clustered by school-subject, are presented in parentheses. The dependent variable is standardized average test scores. Year, semester, and course effects are controlled for in all estimations. *** 1% significant; ** 5% significant; * 10% significant.

Table 3.5: The Effects of Free Semester on Middle School Standardized Test Scores: Math

	(1)	(2)	(3)	(4)
VARIABLES	Standardized Average Test Score for Math			
Treated*Post	-0.053***	-0.053***	-0.053***	-0.050**
	(0.019)	(0.019)	(0.019)	(0.021)
Treated	0.034*	0.050**	0.043**	
	(0.019)	(0.020)	(0.019)	
Private School		0.069*	0.074**	
		(0.036)	(0.037)	
Single Sex School: Male		-0.049	-0.059	
		(0.040)	(0.043)	
Single Sex School: Female		0.050	0.036	
		(0.040)	(0.043)	
School District Effects			✓	
School Effects				✓
Observations	2,262	2,262	2,262	2,262
R-squared	0.004	0.032	0.078	0.591

Notes: Standard errors, clustered by school-subject, are presented in parentheses. The dependent variable is standardized average test scores. Year and semester effects are controlled for in all estimations. *** 1% significant; ** 5% significant; * 10% significant.

Table 3.6: The Effects of Free Semester on Middle School Standardized Test Scores: Reading

	(1)	(2)	(3)	(4)
VARIABLES	Standardized Average Test Score for Reading			
Treated*Post	-0.002	-0.002	-0.002	0.000
	(0.021)	(0.021)	(0.021)	(0.023)
Treated	-0.003	0.005	0.005	
	(0.022)	(0.022)	(0.022)	
Private School		0.028	0.039	
		(0.041)	(0.042)	
Single Sex School: Male		-0.082*	-0.097**	
		(0.042)	(0.046)	
Single Sex School: Female		0.119***	0.104**	
		(0.044)	(0.047)	
School District Effects			✓	
School Effects				✓
Observations	2,262	2,262	2,262	2,262
R-squared	0.000	0.042	0.082	0.572

Notes: Standard errors, clustered by school-subject, are presented in parentheses. The dependent variable is standardized average test scores. Year and semester effects are controlled for in all estimations. *** 1% significant; ** 5% significant; * 10% significant.

Appendix

Table A1: The Spillover Effect of Later School Starting Time on Parents' Sleep by the Number of Teenage Kids

	(1) Both	(2) Mother	(3) Father
VARIABLES		Sleep Duration	
Mean	7hr 14min	7hr 5min	7hr 24min
Treated*Post* One Kid	0.276	13.750	-18.853
	(8.911)	(11.370)	(13.586)
Treated*Post* More Kids	24.625**	29.953**	17.566
	(10.173)	(13.028)	(15.822)
More Than 1 Kid	-2.352	-9.651**	3.536
	(3.224)	(4.120)	(5.037)
Single Parent	-12.868	11.920	-59.658***
	(18.714)	(25.181)	(22.772)
Female	-28.862***		
	(3.617)		
Observations	2,849	1,490	1,359
R-squared	0.066	0.068	0.083

Notes: Robust standard errors are presented in parentheses. The dependent variable is sleep time in minutes. Month, gender, single parent, having more than one kids, having preschooler, age, education, job classification, industry, year, and day effects are controlled for in all estimations. *** 1% significant; ** 5% significant; * 10% significant.

Table A2: The Spillover Effect of Later School Starting Time on Parents' Activities by the Number of Teenage Kids

Panel A: The Spillover Effect of Later School Starting Time on Activities for Mother

Category	Market Work	Non-Sleep Personal Activity	Household Production	Leisure	Other Activities
One Kid	-22.765	5.457	1.878	-4.019	5.699
	(21.242)	(9.354)	(16.922)	(19.099)	(18.992)
More Kids	1.713	14.520	-18.866	-9.944	-17.376
	(23.163)	(10.657)	(19.580)	(20.179)	(17.248)

Panel B: The Spillover Effect of Later School Starting Time on Activities for Father

Category	Market Work	Non-Sleep Personal Activity	Household Production	Leisure	Other Activities
One Kid	11.400	24.588***	1.753	-12.518	-6.369
	(26.159)	(9.224)	(6.621)	(19.558)	(18.037)
More Kids	-4.162	19.326	10.354*	-15.465	-27.619*
	(28.363)	(12.725)	(6.139)	(20.498)	(16.362)

Notes: Robust standard errors are presented in parentheses. The dependent variables are in the number of minutes spent in each activity for mothers (Panel A) and for fathers (Panel B). Month, gender, single parent, having more than one kids, having preschooler, age, education, job classification, industry, year, and day effects are controlled for in all estimations. *** 1% significant; ** 5% significant; * 10% significant.

Table A3: The Effect of Later School Starting Time on Middle School Test Scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Reading and Math			Reading			Math	
VARIABLES					ln (Average Test Score)				
Treated*Post	0.009***	0.009***	0.008***	0.005**	0.005**	0.005**	0.012***	0.012***	0.012***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
8th Grade	-0.009***	-0.009***	-0.009***	0.003**	0.003**	0.003***	-0.021***	-0.021***	-0.020***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
9th Grade	-0.023***	-0.023***	-0.022***	0.001	0.001	0.002	-0.048***	-0.048***	-0.047***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
School District Effects		✓			✓			✓	
School Effects			✓			✓			✓
Observations	33,282	33,282	33,282	16,640	16,640	16,640	16,642	16,642	16,642
R-squared	0.334	0.361	0.586	0.058	0.085	0.497	0.212	0.261	0.632

Notes: Standard errors, clustered by school, are presented in parentheses. The dependent variable is standardized average test scores. Columns (1)-(3) include reading and math test scores along with a subject indicator. Columns (4)-(6) include only reading test scores and Column (7)-(9) include only math test scores. Year, semester, "free semester", and rural effects are controlled for in all estimations. *** 1% significant; ** 5% significant; * 10% significant.

Bibliography

Altonji, Joseph G., Ching-I. Huang, and Christopher R. Taber. 2005. "Estimating the cream skimming effect of private school vouchers on public school students." Unpublished.

Amrein, Audrey L., and David C. Berliner. 2002. "High-stakes testing & student learning." *Education Policy Analysis Archives* 10, no. 18.

Amrein-Beardsley, Audrey, and David C. Berliner. 2003. "Re-analysis of NAEP math and reading scores in states with and without high-stakes tests: Response to Rosenshine." *Education Policy Analysis Archives* 11, no. 25.

Anders, Thomas F., Mary A. Carskadon, William C. Dement, and Kim Harvey. 1978. "Sleep habits of children and the identification of pathologically sleepy children." *Child Psychiatry & Human Development* 9, no. 1: 56-63.

Andrade, Miriam MM, Ana A. Benedito-Silva, Sorahia Domenice, Ivo JP Arnhold, and Luiz Menna-Barreto. 1993. "Sleep characteristics of adolescents: a longitudinal study." *Journal of Adolescent Health* 14, no. 5: 401-406.

Angrist, Joshua D. 2014. "The perils of peer effects." *Labour Economics* 30: 98-108.

Autor, David H. 2003. "Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing." *Journal of Labor Economics* 21, no. 1: 1-42.

Betts, Julian R., and Andrew Zau. 2004. "Peer groups and academic achievement: Panel evidence from administrative data." Unpublished.

Booker, Kevin, Scott M. Gilpatric, Timothy Gronberg, and Dennis Jansen. 2008. "The effect of charter schools on traditional public school students in Texas: Are children who stay behind left behind?." *Journal of Urban Economics* 64, no. 1: 123-145.

Braun, Henry. 2004. "Reconsidering the impact of high-stakes testing." *Education Policy Analysis Archives* 12, no. 1.

Burke, Mary A., and Tim R. Sass. 2013. "Classroom peer effects and student achievement." *Journal of Labor Economics* 31, no. 1: 51-82.

Cardinali, Daniel P. 2008. “Chronoeducation: How the biological clock influences the learning process.” In *The Educated Brain: Essays in Neuroeducation*, ed. Antonio M. Battro, Kurt W. Fischer, and Pierre J. Léna, 110–26. Cambridge, UK: Cambridge University Press.

Cardon, James H. 2003. "Strategic quality choice and charter schools." *Journal of Public Economics* 87, no. 3-4: 729-737.

Carrell, Scott E., Frederick V. Malmstrom, and James E. West. 2008. "Peer effects in academic cheating." *Journal of Human Resources* 43, no. 1: 173-207.

Carrell, Scott E., Richard L. Fullerton, and James E. West. 2009. "Does your cohort matter? Measuring peer effects in college achievement." *Journal of Labor Economics* 27, no. 3: 439-464.

Carrell, Scott E., Teny Maghakian, and James E. West. 2011. "A's from Zzzz's? The causal effect of school start time on the academic achievement of adolescents." *American Economic Journal: Economic Policy* 3, no. 3: 62-81.

Carskadon, Mary A. "Patterns of sleep and sleepiness in adolescents. 1990. "*Pediatrician* 17, no. 1: 5-12.

Carskadon, Mary A., Cecilia Vieira, and Christine Acebo. 1993. "Association between puberty and delayed phase preference." *Sleep* 16, no. 3: 258-262.

Carskadon, Mary A., Christine Acebo, Gary S. Richardson, Barbara A. Tate, and Ronald Seifer. 1997. "An approach to studying circadian rhythms of adolescent humans." *Journal of Biological Rhythms* 12, no. 3: 278-289.

Chiang, Hanley. 2009. "How accountability pressure on failing schools affects student achievement." *Journal of Public Economics* 93, no. 9-10: 1045-1057.

Coleman, James S., Ernest Campbell, Carol Hobson, James McPartland, Alexander Mood, Frederick Weinfeld, and Robert York. 1966. *Equality of Educational Opportunity*. 2 volumes. Washington, D.C.: Office of Education, U. S. Department of Health, Education, and Welfare, U. S. Government Printing Office.

Crowley, Stephanie J., Christine Acebo, and Mary A. Carskadon. 2007. "Sleep, circadian rhythms, and delayed phase in adolescence." *Sleep Medicine* 8, no. 6: 602-612.

Cullen, Julie Berry, Brian A. Jacob, and Steven D. Levitt. 2005. "The impact of school choice on student outcomes: an analysis of the Chicago Public Schools." *Journal of Public Economics* 89, no. 5-6: 729-760.

Deming, David J., Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger. 2014. "School choice, school quality, and postsecondary attainment." *American Economic Review*, 104 (3): 991-1013.

Dills, Angela K. 2005. "Does cream-skimming curdle the milk? A study of peer effects." *Economics of Education Review* 24, no. 1: 19-28.

Eaton, Danice K., Lela R. McKnight-Eily, Richard Lowry, Geraldine S. Perry, Letitia Presley-Cantrell, and Janet B. Croft. 2010. "Prevalence of insufficient, borderline, and optimal hours of sleep among high school students–United States, 2007." *Journal of Adolescent Health* 46, no. 4: 399-401.

Edwards, Finley. 2012. "Early to rise? The effect of daily start times on academic performance." *Economics of Education Review* 31, no. 6: 970-983.

Epple, Dennis, and Richard E. Romano. 1998. "Competition between private and public schools, vouchers, and peer-group effects." *American Economic Review* 88, no. 1: 33-62.

Epple, Dennis, and Richard Romano. 2008. "Educational vouchers and cream skimming." *International Economic Review* 49, no. 4: 1395-1435.

Epple, Dennis, David Figlio, and Richard Romano. 2004. "Competition between private and public schools: testing stratification and pricing predictions." *Journal of Public Economics* 88, no. 7-8: 1215-1245.

Epple, Dennis, Richard E. Romano, and Miguel Urquiola. 2017. "School vouchers: A survey of the economics literature." *Journal of Economic Literature* 55, no. 2: 441-92.

Figlio, David, and Susanna Loeb. 2010. "School accountability." In *Handbook of the Economics of Education*, Vol. 3, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 383-421. Amsterdam: North-Holland.

Figlio, David N., and Cecilia Elena Rouse. 2006. "Do accountability and voucher threats improve low-performing schools?" *Journal of Public Economics* 90, no. 1-2: 239-255.

Figlio, David N., and Joe A. Stone. 2001. "Can public policy affect private school cream skimming?" *Journal of Urban Economics* 49, no. 2: 240-266.

Fogel, Stuart M., and Carlyle T. Smith. 2011. "The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation." *Neuroscience & Biobehavioral Reviews* 35, no. 5: 1154-1165.

Foster, Gigi. 2006. "It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism." *Journal of Public Economics* 90, no. 8-9: 1455-1475.

Frederisksen, Norman. 1994. “The influence of minimum competency tests on teaching and learning.” Educational Testing Service, Princeton NJ: Policy Information Center.

Gais, Steffen, and Jan Born. 2004. "Declarative memory consolidation: mechanisms acting during human sleep." *Learning & Memory* 11, no. 6: 679-685.

Gyeonggi Provincial Office of Education. 2014. “A survey of school start times for all the primary and secondary schools in Gyeonggi Province in 2014 Spring semester” Unpublished Data.

Hansen, Martha, Imke Janssen, Adam Schiff, Phyllis C. Zee, and Margarita L. Dubocovich. 2005. "The impact of school daily schedule on adolescent sleep." *Pediatrics* 115, no. 6: 1555-1561.

Hanushek, Eric A., and Margaret E. Raymond. 2005. "Does school accountability lead to improved student performance?" *Journal of Policy Analysis and Management* 24, no. 2: 297-327.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin. 2003. "Does peer ability affect student achievement?" *Journal of Applied Econometrics* 18, no. 5: 527-544.

Hanushek, Eric A., John F. Kain, Steven G. Rivkin, and Gregory F. Branch. 2007. "Charter school quality and parental decision making with school choice." *Journal of Public Economics* 91, no. 5-6: 823-848.

Heissel, Jennifer, and Samuel Norris. 2017. "Rise and shine: The effect of school start times on academic performance from childhood through puberty." *Journal of Human Resources*: 0815-7346R1.

Hirshkowitz, Max, Kaitlyn Whiton, Steven M. Albert, Cathy Alessi, Oliviero Bruni, Lydia DonCarlos, Nancy Hazen et al. 2015. "National Sleep Foundation's sleep time duration recommendations: methodology and results summary." *Sleep Health* 1, no. 1: 40-43.

Hoxby, Caroline, and Gretchen Weingarth. 2005. "Taking race out of the equation: School reassignment and the structure of peer effects." Unpublished.

Hoxby, Caroline M., and Jonah Rockoff. 2005. "The impact of charter schools on student achievement." Unpublished.

Hsieh, Chang-Tai, and Miguel Urquiola. 2006. "The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program." *Journal of Public Economics* 90, no. 8-9: 1477-1503.

Imberman, Scott A. 2011. "The effect of charter schools on achievement and behavior of public school students." *Journal of Public Economics* 95, no. 7-8: 850-863.

Jacob, Brian A. 2005. "Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools." *Journal of Public Economics* 89, no. 5-6: 761-796.

Jacob, Brian A. 2001. "Getting tough? The impact of high school graduation exams." *Educational Evaluation and Policy Analysis* 23, no. 2: 99-121.

Kirmil-Gray, Kathleen, Jean R. Eagleston, Elizabeth Gibson, and Carl E. Thoresen. 1984. "Sleep disturbance in adolescents: sleep quality, sleep habits, beliefs about sleep, and daytime functioning." *Journal of Youth and Adolescence* 13, no. 5: 375-384.

Klein, Stephen P., Laura Hamilton, Daniel F. McCaffrey, and Brian Stecher. 2000. "What do test scores in Texas tell us?." *Education Policy Analysis Archives* 8: 49.

Korean Educational Statistics Service. 2011-2016. “Changes related to student by city/province administrative district.” Korean Educational Development Institute. https://kess.kedi.re.kr/index (accessed September 2017).

Korean School Information. 2009-2012. “school specific information including high school academic achievement for each course by year, semester, grade, and school level.” Korea Education & Research Information Service. http://www.schoolinfo.go.kr/ (accessed 2012-2014).

Korean School Information. 2013-2015. “school specific information including middle school academic achievement for each course by year, semester, grade, and school level.” Korea Education & Research Information Service. http://www.schoolinfo.go.kr/ (accessed January-March 2016 and March-April 2017).

Ladd, Helen F. 1999. "The Dallas school accountability and incentive program: An evaluation of its impacts on student outcomes." *Economics of Education Review* 18, no. 1: 1-16.

Lankford, Hamilton, and James Wyckoff. 2001. "Who would be left behind by enhanced private school choice?" *Journal of Urban Economics* 50, no. 2: 288-312.

Lufi, Dubi, Orna Tzischinsky, and Stav Hadar. 2011. "Delaying school starting time by one hour: some effects on attention levels in adolescents." *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine* 7, no. 2: 137.

Meijer, A. M., H. T. Habekothe, and G. L. H. Van Den Wittenboer. 2000. "Time in bed, quality of sleep and school functioning of children." *Journal of Sleep Research* 9, no. 2: 145-154.

Nader, Rebecca and Carlyle Smith. 2003. "A role for stage 2 sleep in memory processing." In *Sleep and Brain Plasticity*, edited by Pierre Maquet, Carlyle Smith, and Robert Stickgold, 87-98. Oxford: Oxford University Press.

National Sleep Foundation. 2006. Sleep in America poll: Summary of findings. Arlington, VA: National Sleep Foundation.

Neal, Derek. 2002. "How vouchers could change the market for education." *Journal of Economic Perspectives* 16, no. 4: 25-44.

Richards, Craig E., and Tian Ming Sheu. 1992. "The South Carolina school incentive reward program: A policy analysis." *Economics of Education Review* 11, no. 1: 71-86.

Rockoff, Jonah, and Lesley J. Turner. 2010. "Short-run impacts of accountability on school quality." *American Economic Journal: Economic Policy* 2, no. 4: 119-47.

Rosenshine, Barak. 2003. "High-stakes testing: Another analysis." *Education Policy Analysis Archives* 11: 24.

Rouse, Cecilia Elena, Jane Hannaway, Dan Goldhaber, and David Figlio. 2013. "Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure." *American Economic Journal: Economic Policy* 5, no. 2: 251-81.

Sacerdote, Bruce. 2001. "Peer effects with random assignment: Results for Dartmouth roommates." *The Quarterly Journal of Economics* 116, no. 2: 681-704.

Sadeh, Avi, Reut Gruber, and Amiram Raviv. 2003. "The effects of sleep restriction and extension on school-age children: What a difference an hour makes." *Child Development* 74, no. 2: 444-455.

Selected teachers from Gyeonggi Province secondary and primary schools, telephone survey by the author, September 2017.

Smith, Stephen Samuel, and Roslyn Arlin Mickelson. 2000. "All that glitters is not gold: School reform in Charlotte-Mecklenburg." *Educational Evaluation and Policy Analysis* 22, no. 2: 101-121.

Stickgold, Robert, and Matthew P. Walker. 2007. "Sleep-dependent memory consolidation and reconsolidation." *Sleep Medicine* 8, no. 4: 331-343.

Stinebrickner, Ralph, and Todd R. Stinebrickner. 2006. "What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds." *Journal of Public Economics* 90, no. 8-9: 1435-1454.

Zernike, Kate. 2015. "Obama Administration calls for limits on testing in schools." *The New York Times*, October 24, 2015.

https://www.nytimes.com/2015/10/25/us/obama-administration-calls-for-limits-on-testing-in-schools.html.

Doering, Christopher. 2015. “Obama plan limits standardized testing to no more than 2% of class time.” *USA Today*, October 24, 2015.

https://www.usatoday.com/story/news/2015/10/24/obama-schools-test/74536886/.

Vigdor, Jacob L., and Thomas Nechyba. 2007. “Peer effects in North Carolina public schools.” In *Schools and the Equal Opportunity Problem*, edited by Ludger Woessman and Paul E. Peterson, 73–101. Cambridge, MA: MIT Press.

Walsh, Patrick. 2009. "Effects of school choice on the margin: The cream is already skimmed." *Economics of Education Review* 28, no. 2: 227-236.

Wolfson, Amy R., and Mary A. Carskadon. 1998. "Sleep schedules and daytime functioning in adolescents." *Child Development* 69, no. 4: 875-887.

Zimmerman, David J. 2003. "Peer effects in academic outcomes: Evidence from a natural experiment." *Review of Economics and Statistics* 85, no. 1: 9-23.